



EVALUAR

Revista Evaluar

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

2019

VOL 19 - N°1
ISSN 1667-4545

Validación de la Escala de Actitud hacia la Investigación en Estudiantes Mexicanos de Psicología

Validation Study of the Attitude Scale towards Research among Mexican Psychology Students

Lucía Quezada-Berumen *¹, José Moral de la Rubia¹, René Landero-Hernández¹

¹ - Facultad de Psicología, Universidad Autónoma de Nuevo León, Monterrey, NL., México.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 02/02/2019 Revisado: 21/02/2019 Aceptado: 25/02/2019

Resumen

Esta investigación tiene como objetivo validar la Escala de Actitud hacia la Investigación. Se usó un muestreo de bola de nieve. Participaron 392 estudiantes de psicología de una universidad pública del noreste de México, 74.2% mujeres y 25.8% hombres. El análisis factorial exploratorio arrojó una estructura de tres factores: *afectivo-conductual* de ítems directos o redactados en sentido de aceptación (α ordinal = .90), *cognoscitivo* (α ordinal = .89) y *conductual-afectivo* de ítems inversos o redactados en sentido de rechazo (α ordinal = .87). El ajuste del modelo fue bueno (NFI = .95, RFI = .95, RMSEA = .02 y SRMR = .07) y los tres factores mostraron validez convergente ($AVE > .50$ y $\omega > .80$) y discriminante ($r^2_{F1,F2} < AVE_{F1}$ y AVE_{F2}). La actitud promedio fue favorable hacia la investigación sin diferencia entre mujeres y hombres, salvo por un mayor promedio en el factor *afectivo-conductual* de ítems directos en hombres. El puntaje promedio más alto en estudiantes que piensan dedicarse a la investigación y la correlación positiva entre la escala y la calificación en la materia de investigación proporcionaron evidencias de validez de constructo. Se concluye que la escala presenta consistencia interna y validez de constructo.

Palabras clave: *actitud, investigación, análisis factorial, confiabilidad, México*

Abstract

The aim of this research was to validate the Scale of Attitude towards Research (SAR). A snowball sampling was used. A sample of 392 psychology students, 74.2% women and 25.8% men, from a public university in north-eastern Mexico was collected. The exploratory factor analysis yielded a structure of three factors: *affective-behavioral* factor of positively codified items, phrased in the sense of acceptance (ordinal $\alpha = .90$), *cognitive factor* (ordinal $\alpha = .89$), and *behavioral-affective* factor of negatively codified items, phrased in the sense of rejection (ordinal $\alpha = .87$). The model had a good fit (NFI = .95, RFI = .95, RMSEA = .02, and SRMR = .07) and the three factors showed convergent validity ($AVE > .50$ and $\omega > .80$) and discriminant validity ($r^2_{F1,F2} < AVE_{F1}$ and AVE_{F2}). The average attitude was favorable towards research, with no difference between women and men, except for a greater average in the *affective-behavioral* factor of positive-codified items among men. The highest average on the scale among students who plan to pursue the research area and the positive correlation between the scale and the marks in a research related subject provided evidence of construct validity. We concluded that the scale shows internal consistency and construct validity.

Keywords: *attitude, research, factor analysis, reliability, Mexico*

Introducción

La investigación es una actividad orientada a la búsqueda de nuevos conocimientos en cualquier campo del saber humano. Tradicionalmente, la investigación ha sido una actividad atribuida principalmente a las instituciones de educación superior, con el propósito de responder a las necesidades de la sociedad (De la Cruz, 2013; González-González, Galindo-Miranda, Galindo-Miranda, & Gold-Morgan, 2004). Si bien es cierto que el objetivo fundamental de la educación en pregrado no tiene como función específica formar investigadores, a este nivel educativo sí le corresponde fomentar en sus alumnos actitudes positivas hacia la investigación, de manera que se conviertan en sus usuarios y, en el mejor de los casos, la adopten como una forma habitual de afrontar los problemas que les plantea el ejercicio de su profesión (Aldana, Caraballo, & Babativa, 2016; Ruiz, 2010).

Una de las razones por las que la investigación es importante en los centros docentes es porque estimula el pensamiento crítico y la creatividad. Así, esta constituye un medio por el cual el proceso de aprendizaje se vitaliza y permite combatir el aprendizaje por memorización, el cual tanto ha contribuido a formar profesionales pasivos, sin innovación y con escasa curiosidad e iniciativa personal (Ruiz, 2010). La falta de interés y las actitudes negativas de los estudiantes hacia la ciencia y la tecnología plantean un problema para la educación científica, ya que llevan a conocimientos deficientes sobre la ciencia y una falta de iniciativa e implicación en proyectos científicos, los cuales son necesarios para que el sistema de ciencia y tecnología mantenga su actividad de progreso (Vázquez & Manassero, 2008).

Las actitudes hacia la investigación se pueden considerar tanto causa como efecto, esto es, como determinantes y objetivos del aprendizaje

en materias de investigación. Como causas del aprendizaje, las actitudes positivas lo favorecen y las negativas lo dificultan; como efecto del aprendizaje, las actitudes pueden ser un contenido más de aprendizaje y, por tanto, ser enseñadas y aprendidas en la universidad e instituciones de educación media superior. Una buena actitud hacia el estudio y el aprendizaje en general suele ser garantía de motivación, interés y esfuerzo en los estudiantes. A su vez, los profesores entienden y aprecian las actitudes en este sentido como elementos que favorecen el rendimiento en las materias de investigación. Sin embargo, las actitudes también pueden ser planteadas como una consecuencia o efecto de la educación, es decir, como objetivos a alcanzar por medio de esta (Vázquez-Alonso & Manassero-Mas, 1997).

Hasta el momento se ha hablado de la actitud hacia la investigación científica, pero falta definir este constructo en el cual se centra el presente estudio instrumental. La actitud hacia la investigación científica es una organización duradera y persistente de creencias, sentimientos y disposiciones respecto a la investigación por parte de los individuos de un colectivo (Aldana de Becerra & Joya-Ramírez, 2011). De acuerdo con Estrada-Corona (2012), si se conocen las actitudes manifiestas de una persona hacia el objeto de actitud es posible inferir cuál será su conducta.

Existen algunos instrumentos para evaluar las actitudes hacia la investigación, como la Escala de Actitudes hacia la Investigación desarrollada por Papanastasiou (2005) en Chipre. En Venezuela, se encuentra la Escala de Actitud hacia el Proceso de Investigación Científico-Social en el docente investigador universitario de Blanco y Alvarado (2005). En Colombia, se han desarrollado la Escala de Actitudes hacia la Investigación Científica en Docentes de Metodología de la Investigación por Aldana de Becerra y Joya-Ramírez (2011); y más recientemente la Escala para

medir Actitudes hacia la Investigación (EACIN) en ambientes académicos creada por Aldana et al. (2016).

Con respecto a esta última escala, su validez de contenido ha sido establecida a través de un juicio de expertos y se ha demostrado su confiabilidad a través de la consistencia interna (Aldana et al., 2016). La escala fue diseñada para medir tres aspectos o factores interrelacionados de la actitud: afectivo, cognoscitivo y conductual. El factor *afectivo* refiere a lo que la persona siente y a las emociones que le produce la investigación; el *cognoscitivo*, a lo que la persona sabe o cree saber acerca de la investigación; y el *conductual*, a lo que la persona hace o está dispuesta a hacer con respecto a la investigación (Aldana de Becerra & Joya-Ramírez, 2011). Cabe señalar que este modelo de tres factores correlacionados no ha sido demostrado empíricamente.

Por su versatilidad para evaluar las actitudes hacia la investigación en comunidades académicas, su desarrollo en Latinoamérica y su contenido, la EACIN se considera un instrumento adecuado para medir las actitudes hacia la investigación en estudiantes universitarios mexicanos. Contar con este instrumento facilitaría obtener un mayor conocimiento y un mejor abordaje de la implicación de los estudiantes en actividades científicas dentro de las Instituciones de Educación Superior, las cuales están haciendo un énfasis creciente en la importancia e incorporación de la investigación científica dentro de sus programas educativos. No obstante, sus propiedades métricas no han sido estudiadas en México y el establecimiento de su validez de constructo está en proceso.

Por lo antes mencionado, el propósito de este estudio fue validar la EACIN y medir la actitud hacia la investigación en estudiantes de psicología de una universidad pública mexicana. Se

plantearon seis objetivos: 1) valorar las propiedades de distribución, discriminabilidad y consistencia interna de los ítems, 2) explorar la estructura factorial verificando que sus factores presenten consistencia interna, validez convergente y validez discriminante; 3) describir la distribución de la puntuación total de la EACIN y sus factores; 4) comparar las medias entre los factores e interpretar el nivel actitudinal; 5) evaluar la relación entre la puntuación total de la EACIN y sus factores con la edad, el sexo y el semestre que se cursa y 6) probar la validez de constructo verificando la correlación positiva de la escala con la calificación en la materia de investigación, y el mayor promedio del puntaje en la escala de los estudiantes que se piensa dedicar a la investigación, en comparación con los que se piensan dedicar a otras áreas.

Método

El presente estudio empírico de tipo instrumental se realizó utilizando un diseño ex post facto de corte transversal, ya que no se manipuló ninguna variable y los datos se colectaron en un único punto temporal (Ato, López, & Benavente, 2013).

Participantes

La población de la que se extrajo la muestra fue compuesta por estudiantes de psicología de una universidad pública del estado de Nuevo León, México. Los criterios de inclusión fueron: estar matriculado en el programa de licenciatura en psicología, cursar o haber cursado al menos una materia de investigación (hay una materia en cada semestre) y otorgar el consentimiento informado. Los criterios de exclusión fueron estudiar

la licenciatura en otra facultad u otra universidad y ser estudiante de posgrado. El criterio de eliminación fue tener datos incompletos en la EACIN. Se usó un muestreo no probabilístico guiado por el participante o de bola de nieve y se obtuvo una muestra válida de 392 participantes, lo que representó el 10.5% de la matrícula total de la licenciatura en psicología de la universidad en la que se hizo el estudio (N = 3,749 estudiantes).

De los 392 participantes, 291 (74.2%) fueron mujeres y 101 (25.8%) hombres. La edad osciló entre 17 y 46 años con una media de 19.41 (DE = 2.93). La media de la calificación en la materia de investigación fue 88.24 (DE = 8.07), variando de 40 a 100. El 42.4% de los alumnos cursaban el primer semestre; 9.9% el segundo; 7.1% el tercero; 24.2% el cuarto; 7.4% el quinto y el 9% correspondió a semestres del sexto al décimo. Entre las áreas de trabajo a las que el estudiante piensa dedicarse, dominaron la psicología clínica (54.8%), educativa (17.8%) y laboral (14.8%). Al área de investigación pensaban dedicarse el 2.8% de los participantes.

Instrumentos de medida

Se aplicó un cuestionario en línea integrado por la solicitud de consentimiento informado, dos preguntas sobre información sociodemográfica (sexo y edad), cinco preguntas sobre información académica y la EACIN.

Escala de Actitudes hacia la Investigación (EACIN; Aldana et al., 2016). Consta de 34 ítems tipo Likert con una escala de respuesta de cinco opciones: 0 = *Muy en desacuerdo*, 1 = *En desacuerdo*, 2 = *Ni de acuerdo ni en desacuerdo*, 3 = *De acuerdo* y 4 = *Muy de acuerdo*. Su puntaje se obtiene sumando sus 34 ítems, tras invertir

la puntuación en los ítems 1, 4, 5, 9, 14, 19, 23, 27, 28, 30 y 34, por lo que su rango varía de 0 a 136. Aldana et al. (2016) diseñaron los ítems para medir tres factores correlacionados: *afectivo* con nueve indicadores (ítems 2, 3, 6, 11, 14, 17, 19, 25 y 27), *cognoscitivo* con 12 indicadores (ítems 1, 7, 12, 15, 20, 22, 26, 28, 29, 31, 32 y 33) y *conductual* con 13 indicadores (ítems 4, 5, 8, 9, 10, 13, 16, 18, 21, 23, 24, 30 y 34). Su consistencia interna total es alta (α de Cronbach = .85) y las comunalidades de sus ítems varían de .25 a .61 con una media de .45; no obstante, cinco ítems presentan propiedades de consistencia interna débiles (ítems 1, 6, 10, 18 y 31), al disminuir el valor de la α total con su eliminación (Aldana et al., 2016).

En el presente estudio, se añadió un nuevo ítem inverso, que se ubicó al final de la escala, sobre un aspecto conductual (ítem 35, *Para ser sincero/a realmente lo que menos hago es leer*). Se decidió incluirlo por dos motivos: es un aspecto señalado por los maestros de investigación en las reuniones del área y retoma la observación de que uno de los obstáculos para el aprendizaje de las materias de investigación es que los alumnos no leen (Gómez-Delgado, Bazán-Ramírez, & Villalobos-Galvis, 2017). Los puntajes de la escala y los factores se calcularon sumando los ítems puntuados en dirección hacia actitud positiva, y dividiendo por el número de ítems sumados, con lo que se obtiene un rango continuo de 0 a 4.

Procedimiento

En los salones de clase, maestros del área de investigación invitaron a sus alumnos a que respondieran un cuestionario en línea sobre actitud hacia la investigación. Además, se solicitó a los estudiantes que, una vez terminaran de responder el cuestionario, difundieran el enlace con compa-

ñeros a través de sus redes sociales. El enlace estuvo disponible de octubre a diciembre del 2017. Entre las instrucciones se pidió a los participantes que respondieran la totalidad del cuestionario. La participación en el estudio fue voluntaria (se solicitó un consentimiento informado para participar), no remunerada y anónima. No se registraron datos de identificación personal. Se incluyó un correo electrónico de los responsables de la investigación para resolver cualquier duda o problema suscitado por la misma, por lo que el estudio atendió a las normas de investigación de la [American Psychological Association \(2017\)](#).

Análisis de datos

La discriminabilidad se determinó por la diferencia de tendencia central entre el grupo de puntuaciones altas y bajas en la escala (percentil 25 y 75). El contraste se hizo por la prueba U de Mann-Whitney. Se requirió una diferencia significativa y con valor mayor que .5 (un octavo del rango). Para la consistencia interna se calculó la correlación entre el ítem y el resto de la escala por el coeficiente de correlación poliserial, la comunalidad en la extracción de un factor general por el método de mínimos cuadrados no ponderados desde la matriz de correlación policórica, y el coeficiente alfa ordinal de la escala tras eliminar el ítem. Una correlación menor que .30, una comunalidad menor que .16 y el incremento del valor coeficiente alfa ordinal se interpretaron como indicadores de consistencia interna débil.

El número de factores se determinó por la convergencia del análisis paralelo de Horn (percentil 95 como punto de intersección), coordenadas óptimas (misma especificación) y media mínima de las correlaciones parciales al cuadrado. La extracción de factores se hizo desde la matriz de correlación policórica por el método de míni-

mos cuadrados no ponderados, y la rotación de la matriz factorial por el método promax. Se requirió una carga factorial mínima de .40 en la matriz de configuraciones. Desde la matriz estructural, se calculó la varianza media extraída (AVE por sus siglas en inglés) y el coeficiente omega de McDonald o de confiabilidad compuesta (ω). Una $AVE > .50$ y un coeficiente $\omega \geq .70$ se interpretaron como indicadores de validez convergente ([Borsboom, Mellenbergh, & van Heerden, 2004](#)). Con una $AVE > .40$ se puede hablar de una validez convergente adecuada, si se compensa con una mayor exigencia en la confiabilidad compuesta, como un valor de $\omega > .80$ ([Borsboom et al., 2004](#)). A su vez, una varianza compartida entre dos factores (cuadrado del coeficiente producto-momento de Pearson) menor que .66 y menor que la AVE de cada factor refleja validez discriminante entre estos ([Borsboom et al., 2004](#)). Los valores de consistencia interna de la escala y los factores se calcularon por el coeficiente alfa ordinal; valores $< .50$ se interpretaron como consistencia interna inaceptable, entre .50 y .59 pobre, de .60 a .69 cuestionable, de .70 a .79 aceptable, de .80 a .89 buena y $\geq .90$ excelente ([Gadermann, Guhn, & Zumbo, 2012](#)).

Se estimaron los parámetros del modelo por mínimos cuadrados no ponderados, usando como datos de entrada la matriz de correlación policórica. Los intervalos de confianza al 95% para los estimadores se calcularon por el método percentil con la extracción de 2000 muestras aleatorias. Se evaluó el ajuste del modelo a los datos a través de ocho índices: chi-cuadrada relativa (χ^2/gl), índice de bondad de ajuste de Jöreskog y Sörbom (GFI) y su fórmula corregida (AGFI), índice de ajuste comparativo de Bentler (CFI), índice de ajuste normado de Bentler y Bonett (NFI), índice de ajuste relativo de Bollen (RFI), error cuadrático medio de Steiger y Lindt (RMSEA), y error cuadrático medio (SRMR). Se estipularon como

valores de buen ajuste: $\chi^2/gl \leq 2$, GFI, CFI, NFI y RFI $\geq .95$, AGFI $\geq .90$, RMSEA y SRMR $\leq .05$ y como valores aceptables: $\chi^2/gl \leq 3$, GFI, CFI, NFI y RFI $\geq .90$, AGFI $\geq .85$, RMSEA y SRMR $\leq .08$ (Byrne, 2016).

La normalidad se contrastó por la prueba ómnibus de D'Agostino-Pearson (Poitras, 2006). La comparación de medias entre los factores se hizo por la prueba *t* de Student para muestras emparejadas. El tamaño del efecto se estimó por la *d* de Cohen. Valores de *d* menores que .2 se interpretaron como un tamaño del efecto trivial, entre .2 y .49 pequeño, entre .5 y .79 mediano y mayor que .8 grande (Téllez, García, & Corral-Verdugo, 2015).

Las relaciones de la puntuación total de la EACIN y sus factores con la edad y calificación en la materia de investigación se calcularon por la correlación producto-momento de Pearson y con el semestre con el coeficiente por rangos de Spearman. La relación con el sexo se estableció con una comparación de medias a través de la prueba *t* de Student para muestras independientes. El tamaño del efecto del sexo sobre la actitud hacia la investigación se calculó por la *g* de Hedges con la corrección de sesgo, y este se interpretó como la *d* de Cohen (Téllez et al., 2015).

La relación con el área profesional se estableció con una comparación de medias a través del análisis de varianza. El tamaño del efecto del área profesional sobre la actitud hacia la investigación se calculó por el coeficiente eta al cuadrado. Valores entre .01 y .059 se interpretaron como un tamaño del efecto pequeño, entre .06 y .139 mediano y $\geq .14$ grande (Téllez et al., 2015). Se comparó al grupo del área profesional de investigación con los otros cuatro grupos de áreas profesionales a través de la prueba de Dunnett con un nivel de significación de .05. El tamaño del efecto en estas cuatro comparaciones se determinó con *g*

de Hedges con la corrección de sesgo.

Las correlaciones con la edad y el semestre cursado, ambas variables con leptocurtosis muy fuerte, se calcularon por el coeficiente de correlación por rangos de Spearman (r_s). A través de la correlación por rangos parcial (r_p) se calculó la correlación entre la escala y la edad, parcializando en ambas variables la varianza del semestre cursado, es decir, sustrayendo la covarianza que el semestre cursado tiene con ambas variables (Conover, 1999). Las correlaciones con la calificación en la materia de investigación se calcularon por el coeficiente de correlación producto-momento de Pearson y su intervalo de confianza se determinó por el método de muestreo repetitivo con la extracción de 2,000 muestras aleatorias (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2014). Valores absolutos de r o r_s entre .01 y .029 se interpretaron como una fuerza de la asociación débil, entre .30 y .499 moderada, entre .50 y .699 fuerte y mayor o igual que .70 muy fuerte (Téllez et al., 2015).

Los cálculos estadísticos se realizaron en SPSS versión 24 (IBM, 2016) con el módulo R 4.1, Excel 2007 y AMOS versión 16. Se usó un nivel de significación de .05 con un contraste a dos colas, salvo en la prueba de normalidad, en la que el contraste fue a una cola.

Resultados

Distribución, discriminabilidad y consistencia interna de los ítems

Ningún ítem presentó efecto suelo o techo. El porcentaje más alto en las categorías primera y última de respuestas fue de 49.5% ($< 80\%$). No obstante, los ítems 12 y 34 mostraron asimetría positiva extrema, coeficiente de asimetría percentílico (CAP) = 1 y el ítem 29 asimetría negativa

extrema, $CAP = -1$; a su vez, los ítems 12, 16 y 29 presentaron una leptocurtosis (colas adelgazadas) extrema, curtosis percentilica corregida (CPC) = .24. El ítem 26 tuvo una variabilidad mínima, rango semi-intercuartílico (RSI) = 0, y un aplanaamiento (colas engrosadas) extremo, $CPC = -.26$. Cabe señalar que ninguno de estos cinco ítems presentó problemas de discriminabilidad o consistencia interna, por lo que no fueron eliminados (Tabla 1).

Los ítems 28 y 32, correspondientes al factor cognoscitivo, tuvieron una diferencia media menor que .05 entre los grupos de puntuaciones altas y bajas en la escala, y mostraron una consistencia interna débil, ya que su correlación con el resto de la escala fue menor que .30, su comunalidad menor que .16 y su eliminación incrementó el coeficiente alfa ordinal de la escala, α ordinal $> .94$ (Tabla 1), por lo que ambos ítems no fueron incluidos en los subsiguientes análisis. La consistencia interna de la escala calculada por el coeficiente alfa de Cronbach también fue muy alta ($\alpha = .92$).

Exploración de la estructura factorial

El análisis paralelo de Horn, coordenadas óptimas y la media mínima de las correlaciones parciales al cuadrado y a la cuarta potencia convergieron en que el número de factores sustantivos es cuatro. Tras extraerse cuatro factores se explicó el 48.7% de la varianza de los ítems. El primer factor quedó conformado por doce ítems (ítems 2, 3, 6, 8, 10, 11, 13, 16, 17, 21, 24 y 25), el segundo por nueve indicadores (ítems 7, 12, 15, 20, 22, 26, 29, 31 y 33), el tercero por dos (ítems 18 y 30) y el cuarto por diez (ítems 1, 4, 5, 9, 14, 19, 23, 27, 34 y 35). Cabe mencionar que el tercer factor tuvo un número insuficiente de indicadores (ítems 18 y 30) y no mostró un verdadero contenido teórico (pues sus dos ítems son repetitivos

en contenido). Por lo anterior y al mostrar menor consistencia interna en la escala, se optó por eliminar el ítem 30 (Tabla 1).

Al extraer tres factores con los restantes 32 ítems, se explicó el 45.3% de la varianza. Tras la rotación, se definió un primer factor con trece indicadores (ítems 2, 3, 6, 8, 10, 11, 13, 16, 17, 18, 21, 24 y 25), que corresponden a ítems directos de aspectos *afectivos* y *conductuales* (ACID). Su consistencia interna fue excelente (α ordinal = .90) y una validez convergente adecuada (AVE = .42; $\omega = .90$). El segundo factor quedó formado por nueve indicadores (ítems 7, 12, 15, 20, 22, 26, 29, 31 y 33) que corresponden a ítems del factor *cognoscitivo* (COG). Su consistencia interna fue buena (α ordinal = .89) y con adecuada validez convergente (AVE = .49; $\omega = .90$). El tercer factor quedó integrado por diez indicadores (ítems 1, 4, 5, 9, 14, 19, 23, 27, 34 y 35), que corresponden a ítems inversos sobre aspectos *conductuales* y *afectivos* (CAII). Su consistencia interna fue buena (α ordinal = .87), su varianza media extraída fue próxima a .40 (AVE = .39) y su confiabilidad compuesta fue mayor que .80 ($\omega = .86$). La varianza compartida entre pares de factores varió de .29 a .38 con una media de .32. En todos los casos, esta varianza compartida entre pares de factores fue menor que la varianza media extraída de cada factor, por lo que los tres factores presentaron validez discriminante.

Para complementar el análisis exploratorio, se especificó un modelo de tres factores correlacionados con sus residuos independientes (Figura 1) usando la muestra de 392 estudiantes en que se aplicó el análisis factorial exploratorio. Todas las estimaciones por intervalo fueron significativas. Los tres factores presentaron validez convergente (AVE = .64 y $\omega = .90$ para el factor ACID, AVE = .69 y $\omega = .89$ para el factor COG y AVE = .63 y $\omega = .87$ para el factor CAII). La varianza compartida entre los factores varió de .38 a .52 ($< .66$) con una media de .43 ($< .50$),

Tabla 1

Distribución, discriminabilidad y consistencia interna de los 35 ítems de la EACIN.

Ítem	Porcentaje		Estadísticos descriptivos				Discriminabilidad		Consistencia		
	MD	MA	Mdn	RSI	CAP	CPC	DM	Z _U	r _{PS}	λ ²	α ord _{ti}
1	49.5	1.5	1	0.5	0	-.013	-1.239	-9.941***	-.559	.353	.934
2	3.1	7.9	2	0.5	0	-.013	0.811	-6.369***	.445	.201	.935
3	5.4	7.4	2	0.5	0	-.013	1.318	-8.385***	.522	.325	.935
4	21.9	0.3	1	0.5	0	-.013	-1.043	-7.796***	-.474	.251	.935
5	49	0.5	1	0.5	0	-.013	-1.195		-.657	.463	.933
6	5.1	11.7	2	0.5	.3	-.096	1.589	-9.851***	.632	.435	.934
7	0.3	37.8	3	0.5	0	-.013	1.115	-9.277***	.568	.377	.934
8	0.3	28.3	3	0.5	0	-.013	1.152	-9.070***	.548	.335	.934
9	7.1	2.8	2	0.5	0	-.013	-1.049	-7.316***	-.476	.233	.935
10	1.5	15.8	3	0.5	0	-.013	1.286	-9.564***	.603	.386	.934
11	1.3	11	2	0.5	0	-.096	1.378	-9.821***	.654	.457	.933
12	0.3	47.2	3	0.5	1	.237	0.952	-9.418***	.534	.324	.934
13	1.8	13.8	2	0.5	.3	-.096	1.180	-7.734***	.509	.253	.935
14	8.2	1.5	2	0.5	0	-.013	-1.128	-8.436***	-.540	.317	.934
15	0.8	47.2	3	0.5	.2	.031	0.806	-8.094***	.443	.244	.935
16	4.6	7.4	2	1	0	.237	1.473	-8.915***	.614	.430	.934
17	5.1	9.4	2	0.5	0	-.013	1.771		.741	.607	.932
18	1.8	10.7	2	0.5	.3	-.096	1.116	-7.888***	.506	.268	.935
19	15.3	2	1	0.5	0	-.013	-1.415	-9.743***	-.625	.444	.933
20	0.5	30.6	3	0.5	0	-.013	1.013	-8.670***	.529	.322	.934
21	0.8	10.2	3	0.5	-.3	-.096	1.117	-8.465***	.506	.287	.935
22	0.3	32.1	3	0.5	0	-.013	1.093	-9.882***	.614	.422	.933
23	17.1	2.3	1	0.5	.3	-.096	-1.185	-7.844***	-.423	.167	.936
24	7.7	4.6	2	0.5	0	-.013	1.086	-6.935***	.479	.228	.935
25	2.8	5.9	2	0.5	0	-.013	1.220	-8.310***	.572	.349	.934
26	0.3	21.9	3	0	0	-.263	1.132	-9.204	.586	.389	.934
27	12.8	1.8	1	0.5	0	-.096	-1.459	-9.745***	-.634	.426	.934
28	2.8	21.9	3	0.5	-.3	-.096	0.209	-2.370*	.192	.047	.937
29	0.5	58.7	4	0.5	-1	.237	0.845	-8.619***	.524	.324	.934
30	13	2	1	0.5	.3	-.096	-1.165	-8.246***	-.444	.188	.935
31	0.5	37	3	0.5	0	-.013	1.022	-9.018***	.494	.284	.934
32	3.3	13.5	2	0.5	.3	-.096	0.492	-3.293***	.196	.043	.938
33	1	26.5	3	0.5	0	-.013	0.890	-7.472***	.445	.216	.935
34	8.9	2.8	1	0.5	1	-.013	-1.073	-7.854***	-.477	.234	.935
35	30.4	2.6	1	1	.3	.070	-1.327	-8.297***	-.407	.173	.936

Nota. Porcentaje de casos para la primera opción de respuesta (MD = muy en desacuerdo) y para la última opción (MA = muy de acuerdo). Estadísticos descriptivos: Mdn = mediana, RSI = rango semiintercuartílico, CAP = coeficiente de asimetría percentílico de Kelly, CPC = coeficiente de curtosis percentílico corregido. Discriminabilidad: DM = diferencia media en el ítem entre los grupos de puntuaciones altas y bajas en la escala (suma de ítems puntuados en sentido de actitud favorable), Z_U = valor del estadístico U de Mann-Whitney al comparar a los grupos de puntuaciones altas y bajas, significación de Z_U en un contraste a dos colas: ns (no significativo) $p > .05$, * $p \leq .05$, ** $p \leq .01$ *** $p \leq .001$. Consistencia interna: r_{PS} = correlación poliserial entre el ítem y el resto de la escala, λ² = comunalidad en la extracción de un factor por mínimos cuadrados no ponderados desde la matriz de correlación policórica, y α ord_{ti} = coeficiente alfa ordinal de la escala tras eliminar el ítem.

siendo la varianza compartida entre cada par de factores menor que la varianza media extraída de cada factor, por lo que los tres factores presentaron validez discriminante. La bondad de ajuste fue buena por siete índices, $\chi^2/gl = 566.68/461 = 1.23 < 2$; GFI = .965, AGFI = .960, CFI = .991, NFI = .954 y RFI = .951 > .95; y RMSEA = .024 < .05, y aceptable por uno, SRMR = .072 < .075.

Distribución de la puntuación total de la EACIN-32 y sus tres factores

Las distribuciones de la puntuación total (DP = 4.21, $p = .122$) y CAII (DP = .18, $p = .912$) se ajustaron a una curva normal por la prueba de D'Agostino-Pearson. La distribución del factor ACID fue simétrica y unimodal, con una media (M = 2.32) y mediana (Mdn = 2.31) muy próxi-

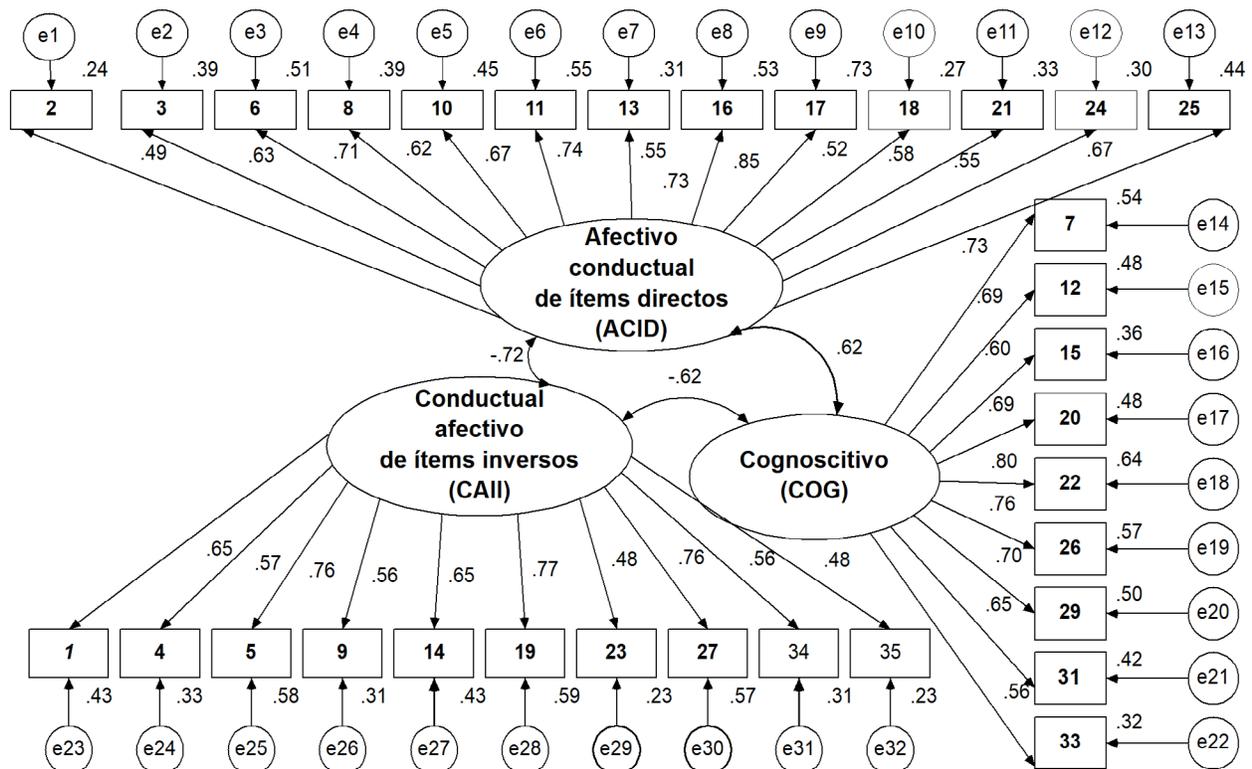
mas. Aunque presentó ligera leptocurtosis (K = .66, IC del 95%: .18, 1.15), la hipótesis nula de distribución normal se pudo sostener por la prueba de D'Agostino-Pearson con un nivel de significación de .01. (DP = 8.30, $p = .016$). La distribución del factor COG presentó concentración en los valores altos (K = 1.29, IC del 95%: .80, 1.77) con colas asimétricas y sesgo hacia los valores por debajo de la mediana (Sk = -.60, IC del 95%: -.84, -.36), lo que no se ajusta a una curva normal (DP = 50.81, $p < .001$).

Comparación de medias entre los factores

Debido al ajuste a la normalidad de la distribución en la puntuación total de la EACIN-32 y el factor CAII, una buena aproximación a la normalidad en el caso del factor ACID y

Figura 1

Modelo de tres factores correlacionados estimado por mínimos cuadrados no ponderados desde la matriz de correlación policórica.



asimetría y curtosis leves en el factor COG, se hizo la comparación global de las medias a través del análisis de varianza de medidas repetidas. La hipótesis nula de esfericidad o independencia entre los residuos se mantuvo por la prueba de Mauchly: $W = .99$, $\chi^2_{(2, N = 392)} = 2.26$, $p = .345$. Hubo diferencia significativa entre las medias de los tres factores ($F_{[2, 324]} = 546.71$, $p < .001$) con una potencia unitaria y un tamaño del efecto muy grande, $\eta^2 = .58$. Al hacer las comparaciones por pares a posteriori con la prueba t de Student para muestras emparejadas con la corrección de Bonferroni sobre el nivel de significación ($2*\alpha/[k*(k-1)] = .1/6 = .017$), la media del factor cognoscitivo fue significativamente mayor que la del factor ACID ($t_{[391]} = 32.65$, $p < .001$) y la del factor CAII ($t_{[391]} = 19.11$, $p < .001$); en ambos casos con un tamaño del efecto grande (d de Cohen = 1.65 y .97, respectivamente). A su vez, la media del factor ACID fue significativamente mayor que la media del factor CAII ($t_{[391]} = 13.68$, $p < .001$) con un tamaño del efecto moderado (d de Cohen = .69).

Relación con las variables sociodemográficas y académicas

Al comparar las medias por la prueba t de Student, en los cuatro casos, se pudo asumir la hipótesis de igualdad de varianzas por la prueba de Levene. La media de los 101 hombres ($M = 2.47$, IC del 95%: 2.35, 2.59) fue significativamente mayor que la media de las 291 mujeres ($M = 2.27$, IC del 95%: 2.20, 2.33) en el factor ACID ($t_{[390]} = 3.07$, $p = .002$). El tamaño del efecto del sexo sobre el factor ACID fue pequeño (g de Hedges = .35, IC del 95%: .13, .58). Las medias entre hombres y mujeres fueron estadísticamente equivalentes en la puntuación total de EACIN-32 ($t_{[390]} = 1.77$, $p = .078$), el factor COG ($t_{[390]} = -.52$,

$p = .605$) y el factor CAII ($t_{[390]} = .96$, $p = .339$).

Las comparaciones de medias en la EACIN-32 y sus tres factores por el área profesional se hicieron por el análisis de varianza de una vía para grupos independientes. Sólo se compararon cinco grupos: área clínica, laboral, educativa, social y de investigación. Se excluyeron los casos de indecisos, o que habían señalado neuropsicología u otra opción por su baja frecuencia. Al mantenerse la hipótesis nula de igualdad de varianza entre los cinco grupos por la prueba de Levene ($F_{[4, 367]} = .25$, $p = .910$ para EACIN-32; $F_{[4, 367]} = .20$, $p = .939$ para ACID; $F_{[4, 367]} = .21$, $p = .930$ para COG, $F_{[4, 367]} = .28$, $p = .894$ para CAII), las comparaciones se realizaron con la prueba de Dunnett.

El área de trabajo a la que el estudiante piensa dedicarse tras la titulación tuvo un efecto significativo y de tamaño mediano sobre la puntuación total de la EACIN-32 ($\eta^2 = .093$) y sobre el factor ACID ($\eta^2 = .102$); fue de tamaño pequeño sobre el factor CAII ($\eta^2 = .055$) y el factor COG ($\eta^2 = .036$). Las cuatro comparaciones entre los grupos del área clínica, laboral, educativa y social con el grupo del área de investigación fueron significativas en la puntuación total ($p \leq .05$) y el factor ACID. En el factor CAII, estas comparaciones fueron significativas ($p \leq .05$) a excepción del grupo del área social y el de investigación ($p = .124$). En el factor COG, sólo la comparación entre el grupo del área laboral y de investigación fue significativa ($p \leq .01$). En todas las comparaciones, la media del grupo del área de investigación fue más alta. La media más baja fue la del área laboral tanto en la puntuación total como en los tres factores. El tamaño del efecto, al ser estimado por la g de Hedges corregida de sesgo, fue grande en todas las comparaciones significativas (g de Hedges = - 0.94 a -1.84).

La puntuación total en la EACIN-32 y los factores ACID y CAII tuvieron correlaciones significativas, positivas y con una fuerza de aso-

ciación débil con la calificación en la materia de investigación, la edad y el semestre cursado. El factor COG presentó correlación positiva y con una fuerza de asociación débil con la calificación promedio, pero fue independiente de las dos variables sociodemográficas. Las correlaciones de la edad con la EACIN-32 y los factores ACID y CAII no fueron significativas al parcializar la varianza del semestre cursado (Tabla 2).

Discusión

Se formuló como primer objetivo valorar si los 35 ítems de la EACIN presentaban problemas de efecto techo o suelo, falta de discriminabilidad y consistencia interna débil. Por los resultados del estudio original de Aldana et al. (2016) con estudiantes y maestros colombianos, se esperaba que los ítems 1, 6, 10, 18 y 31 mostraran una consistencia interna débil. En la presente muestra, ninguno de estos cinco ítems mostró problemas de consistencia interna, discriminabilidad o distribución con excesiva asimetría o curtosis; tampoco el nuevo ítem (ítem 35) sobre escaso tiempo dedicado a lectura presentó problemas. Aldana et

al. (2016) con los 34 ítems obtuvieron una consistencia interna total alta por el coeficiente alfa de Cronbach. En el presente estudio la consistencia interna total fue mayor: con 35 o 34 ítems fue muy alta por el coeficiente alfa ordinal e incluso por el coeficiente de Cronbach.

Se estableció como segundo objetivo explorar la estructura factorial y verificar si los factores cumplían con criterios de consistencia interna, validez convergente y validez discriminante. Se esperaba una estructura de tres factores por los campos de contenido definidos a la hora de diseñar y seleccionar los ítems: afectivo, cognoscitivo y conductual (Aldana et al., 2016). Además, al formar parte de un mismo constructo, se esperaba que los tres factores estuvieran correlacionados y que cumplieran con criterios de consistencia interna, validez convergente y validez divergente. No se reprodujo la estructura esperada, pero sí una afin. El número de factores obtenidos fue tres. Los contenidos sobre aspectos afectivos y conductuales se unieron para diferenciarse por un aspecto formal de la dirección de la redacción de sus ítems: directa o inversa (ACID y CAII). El tercer factor se definió con los ítems sobre aspectos cognoscitivos (COG). Los tres factores

Tabla 2

Correlaciones con semestre cursado, edad y calificación en la materia de investigación.

Escala y sus factores	Semestre	Edad	Edad(sem)	Calificación
	r_s	r_s	r_p	R
EACIN-32	.110* (.011, .209)	.129* (.030, .228)	.075 ^{ns} (-.025, .174)	.177*** (.078, .270)
ACID	.103* (.004, .202)	.130** (.031, .229)	.082 ^{ns} (-.017, .182)	.156** (.056, .253)
COG	.043 ^{ns} (-.056, .142)	.051 ^{ns} (-.048, .150)	.030 ^{ns} (-.070, .129)	.109* (.010, .202)
CAII	.111* (.012, .210)	.115* (.016, .214)	.054 ^{ns} (-.045, .154)	.169*** (.067, .268)

Nota. r_s = coeficiente de correlación por rangos de Spearman, r_p = coeficiente por rangos parcial, r = coeficiente de correlación producto-momento de Pearson. Intervalo de confianza fijado en 95%. Significación en un contraste a dos colas: ns (no significativa) $p > .05$, * $p \leq .01$, ** $p \leq .001$, *** $p \leq .001$.

correlacionaron, pero con buena validez discriminante. El factor ACID mostró una consistencia interna muy alta y los otros dos (COG y CAII) alta. Al especificarse cada factor con sus indicadores exactos, los tres factores claramente mostraron validez convergente y el ajuste del modelo fue bueno. Por tanto, la escala muestra validez de constructo estructural, y este modelo de tres factores se puede adoptar como hipótesis estructural para futuros análisis confirmatorios con muestras independientes. El presente estudio constituye una investigación exploratoria, ya que el análisis factorial confirmatorio se realizó en la misma muestra que el exploratorio (Byrne, 2016).

El tercer objetivo enunciado fue describir la distribución de la puntuación total de la EACIN-32 y sus factores. Muy vinculado a este se encuentra el cuarto objetivo sobre interpretar el nivel actitudinal promedio y comparar los promedios de los factores. La expectativa era que el nivel actitudinal fuese alto y muy alto con una distribución con sesgo hacia los valores altos, debido a la política y programas docentes de las facultades públicas que fomentan la investigación y una ciencia basada en evidencias (Aldana et al., 2016; Blanco & Alvarado, 2005; De la Cruz, 2013; Pelcastre-Villafuerte, Gómez-Serrato, & Zavala, 2015).

Se encontró ajuste a la distribución normal en las puntuaciones del factor CAII y en la puntuación total de la escala (EACIN-32) y una buena aproximación a la normalidad en el factor ACID con simetría y leve leptocurtosis. El factor COG mostró colas asimétricas, con su cola más larga hacia la izquierda, menor variabilidad relativa y una mayor concentración de lo esperado para una curva normal en los valores altos; no obstante, tanto la asimetría negativa como la leptocurtosis fueron leves y su perfil fue acampanado. Por tanto, esta escala de actitud y sus factores parecen

presentar una distribución propia de rasgos adaptativos, en los que la mayoría (68%) se concentra en torno a una desviación estándar de la media (expresión promedio o normal), una minoría se desvía por exceso (16%) y otra minoría por defecto (16%) de esta expresión media o con mayor éxito adaptativo (Lyon, 2012).

Si se divide el rango continuo de las puntuaciones en la escala y sus tres factores, de 0 a 4, en cinco intervalos de amplitud constante ($[\text{valor máximo} - \text{valor mínimo}]/\text{número de ítems} = [4 - 0]/5 = .8$) en correspondencia con los cinco valores ordinales de respuesta a los ítems, se puede interpretar el nivel actitudinal. Valores entre 0 y .79 reflejan una actitud muy desfavorable hacia la investigación, entre .8 y 1.59 desfavorable, entre 1.6 y 2.39 neutra, entre 2.4 y 3.19 favorable y entre 3.2 y 4 muy favorable. Siguiendo este planteamiento de interpretación, la media en el factor COG ($M = 3.20$) refleja una actitud promedio muy favorable, las medias en el total de escala y el factor ACID ($M = 2.68$ y 2.67 , respectivamente) reflejan actitudes favorables y la media en el factor CAII ($M = 2.32$) refleja una actitud neutra. Considerando que la media fue diferencial entre los tres factores, los aspectos cognitivos o intelectuales de la actitud hacia la investigación están más definidos que los aspectos prácticos o de conducta, lo cual es un fenómeno usual en valores transmitidos en ámbitos escolares y académicos, como ocurre con las conductas de salud sexual (Moral & Ortega, 2009) y hábitos saludables (Brazil, Milagres, Duarte, & Rocha, 2018).

El quinto objetivo consistió en evaluar la relación de la EACIN-32 con la edad y el sexo. Conforme con la expectativa, el promedio en la puntuación total y dos de los factores fue equivalente entre ambos sexos (Pelcastre-Villafuerte et al., 2015). Sólo hubo una diferencia significativa con un tamaño del efecto pequeño con un mayor

promedio en el factor ACID en hombres que en mujeres. Precisamente, la lógica analítica e instrumental de la ciencia es más afín al rol de género masculino que al femenino. Probablemente, si se midiese y controlase el rol de género, la correlación resultaría espuria (Kumari & Saraladevi, 2014). También, conforme con la expectativa se obtuvo una correlación positiva con la edad. Esta relación puede atribuirse a un mayor tiempo de exposición a los valores científicos inculcados por los maestros (Estrada-Corona, 2012; Vázquez & Manassero, 2008), apoya esta interpretación el hecho de que la correlación dejó de ser significativa una vez se parcializó el semestre cursado. No obstante, al ser mayor la correlación con la edad que con el semestre, el hecho de tener una mayor madurez para la reflexión y adopción de dichos valores podría ser también un factor contribuyente.

El sexto objetivo fue aportar una prueba de validez de constructo al verificar si la escala presenta un mayor promedio entre los estudiantes que piensan dedicarse a la investigación y si tiene una correlación positiva con la calificación en la materia de investigación. Se cumplieron ambas hipótesis, lo que aporta evidencias de validez de constructo de tipo convergente a la escala. El tamaño del efecto del área de investigación sobre la puntuación total de la EACIN-32 fue mediano en la comparación total, por lo que resulta grande en las comparaciones por pares. La fuerza de asociación entre la puntuación de la EACIN-32 y la calificación promedio fue débil, por lo que probablemente otras variables como hábitos de estudio, promedio en las materias e inteligencia tengan asociación más fuerte.

Una primera limitación del estudio fue el uso de un muestreo no probabilístico. Por lo tanto, las inferencias con la debida precaución se limitan a la facultad de psicología en la que se recolectó la muestra. Aunque debe tenerse en cuenta que la

tasa de casos fue grande, cubre más de un décimo de su población. Una segunda limitación fue el diseño ex post facto de corte transversal, que impide interpretar las asociaciones halladas en términos causales. Una tercera limitación fue que la actitud hacia la investigación se evaluó a través de un solo instrumento de autoinforme, por lo que las conclusiones se circunscriben a esta escala.

Se recomienda el uso de la EACIN-32 para la evaluación y estudio de la actitud hacia la investigación en contextos universitarios. Dado que la puntuación de la escala presenta un buen ajuste a la normalidad, se sugiere contrastar el modelo de tres factores correlacionados propuesto con una muestra probabilística y a su vez estimar los baremos de la escala, los cuales se pueden establecer a través de puntuaciones *T*. El muestreo se podría hacer con estudiantes universitarios. Se podrían aportar nuevas pruebas de validez de constructo concurrentes al verificar la correlación fuerte o muy fuerte de la EACIN-32 con otras escalas de actitud hacia la investigación, como las de Papanastasiou (2005) y Blanco y Alvarado (2005). De igual manera, se podría estudiar la relación de la EACIN-32 con la deseabilidad social. Se esperaría que el efecto de la deseabilidad social sea trivial o débil, y que esta tenga más efecto sobre el factor cognoscitivo, al ser más afín a los contenidos inculcados en las aulas y laboratorios por los maestros.

Referencias

- Aldana, G. M., Caraballo, G. J., & Babativa, D. A. (2016). Escala para medir Actitudes hacia la Investigación (EACIN): Validación de contenido y confiabilidad. *Aletheia. Revista de Desarrollo Humano, Educativo y Social Contemporáneo*, 8(2), 104-121. Recuperado de <http://aletheia.cinde.org.co/index.php/ALETHEIA/article/view/325/253>

- Aldana de Becerra, G. M., & Joya-Ramírez, N. S. (2011). Actitudes hacia la investigación en docentes de metodología de la investigación. *Tabula Rasa*, *14*, 295-309. doi: [10.25058/20112742.428](https://doi.org/10.25058/20112742.428)
- American Psychological Association (2017). *Ethical principles of psychologists and code of conduct. With the 2016 amendment to standard 3.04*. Washington, DC: American Psychological Association. Recuperado de <https://www.apa.org/index>
- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, *29*(3), 1038-1059. doi: [10.6018/analesps.29.3.178511](https://doi.org/10.6018/analesps.29.3.178511)
- Blanco, N., & Alvarado, M. E. (2005). Escala de actitud hacia el proceso de investigación científico social. *Revista de Ciencias Sociales*, *11*(3), 537-544. Recuperado de <http://produccioncientificaluz.org/index.php/rcs/index>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061-1071. doi: [10.1037/0033-295X.111.4.1061](https://doi.org/10.1037/0033-295X.111.4.1061)
- Brazil, J. M., Milagres, M. P., Duarte, A. C. S., & Rocha, R. M. (2018). Social representations of university students on healthy habits. *Journal of Nursing UFPE on line*, *12*(1), 189-193. doi: [10.5205/1981-8963-v12i1a23502p189-193-2018](https://doi.org/10.5205/1981-8963-v12i1a23502p189-193-2018)
- Byrne, B. M. (2016). *Structural equation modelling with amos. Basic concepts, applications, and programming* (3ª. ed.). New York, NY: Routledge. doi: [10.4324/9781315757421](https://doi.org/10.4324/9781315757421)
- Conover, W. J. (1999). *Practical nonparametric statistics* (3ª. ed.). New York, NY: John Wiley & Sons.
- De la Cruz, C. (2013). Actitudes hacia la investigación científica en estudiantes universitarios: Análisis en dos universidades nacionales de Lima. *PsiqueMag*, *2*(1), 1-16. Recuperado de <http://ojs.ucvlima.edu.pe/index.php/psiquemag/article/view/12/12>
- Estrada-Corona, A. (2012). La actitud del individuo y su interacción con la sociedad. Entrevista con la Dra. María Teresa Esquivias Serrano. *Revista Digital Universitaria*, *13*(7), 1-12. Recuperado de <http://www.revista.unam.mx>
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1-13. Recuperado de <http://pa-reonline.net>
- Gómez-Delgado, Y. A., Bazán-Ramírez, A., & Villalobos-Galvis, F. H. (2017). Factores del estudiante que dificultan la enseñanza-aprendizaje de competencias de investigación. *Interacciones. Revista de Avances En Psicología*, *3*(3), 101-110. doi: [10.24016/2017.v3n3.68](https://doi.org/10.24016/2017.v3n3.68)
- González-González, J., Galindo-Miranda, N. E., Galindo-Miranda, J. L., & Gold-Morgan, M. (2004). *Los paradigmas de la calidad educativa. De la autoevaluación a la acreditación*. Ciudad de México: Unión de Universidades de América Latina.
- Hesterberg, T. C., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (2014). Bootstrap methods and permutation tests. En D. S. Moore, G. McCabe & B. A. Craig (Eds.), *Introduction to the practice of statistics* (8ª. ed.). New York, NY: W.H. Freeman and Company.
- IBM Corporation (2016). IBM SPSS Statistics for Windows, Version 24.0 [software de cómputo]. Armonk, NY: IBM.
- Kumari, S. N., & Saraladevi, K. (2014). Gender differences in attitude towards science teaching and learning styles among adolescents. *International Journal of Science and Research*, *3*(10), 50-58. Recuperado de <https://www.ijsr.net>
- Lyon, A. (2012). Mathematical explanations of empirical facts, and mathematical realism. *Australasian Journal of Philosophy*, *90*(3), 559-578. doi: [10.1080/00048402.2011.596216](https://doi.org/10.1080/00048402.2011.596216)
- Moral, J., & Ortega, M. E. (2009). Representación social de la sexualidad y actitudes en estudiantes universitarios mexicanos. *Revista de Psicología Social*, *24*(1), 65-79. doi: [10.1174/021347409786922998](https://doi.org/10.1174/021347409786922998)

- Papanastasiou, H. C. (2005). Factor structure of the “Attitudes Toward Research” Scale. *Statistics Education Research Journal*, 4(1), 16-26. Recuperado de <https://iase-web.org/Publications.php?p=SERJ>
- Pelcastre-Villafuerte, L., Gómez-Serrato, A. R., & Zavala, G. (2015). Actitudes hacia la ciencia de estudiantes de educación preuniversitaria del centro de México. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 12(3), 475-490. doi: [10.25267/rev_eureka_ensen_divulg_cienc.2015.v12.i3.06](https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2015.v12.i3.06)
- Poitras, G. (2006). More on the correct use of omnibus tests for normality. *Economics Letters*, 90(3), 304-309. doi: [10.1016/j.econlet.2005.08.016](https://doi.org/10.1016/j.econlet.2005.08.016)
- Ruiz, J. (2010). Importancia de la investigación. *Revista Científica*, 20(2), 125-126. Recuperado de <https://www.redalyc.org/articulo.oa?id=95912322001>
- Téllez, A., García, C. H., & Corral-Verdugo, V. (2015). Effect size, confidence intervals and statistical power in psychological research. *Psychology in Russia: State of the Art*, 8(3), 27-46. doi: [10.11621/pir.2015.0303](https://doi.org/10.11621/pir.2015.0303)
- Vázquez-Alonso, A., & Manassero-Mas, M. A. (1997). Una evaluación de las actitudes relacionadas con la ciencia. *Enseñanza de las Ciencias*, 15(2), 199-213. Recuperado de <http://www.raco.cat/index.php/Ensenanza>
- Vázquez, A., & Manassero, M. A. (2008). El declive de las actitudes hacia la ciencia de los estudiantes: Un indicador inquietante para la educación científica. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 5(3), 274-292. doi: [10.25267/Rev_Eureka_ensen_divulg_cienc.2008.v5.i3.03](https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2008.v5.i3.03)

Anexo. EACIN-32

A continuación encontrarás una serie de afirmaciones relacionadas con la investigación (clases de investigación, análisis de datos, etc.), por favor marque con una X la respuesta con la cual te sientas más identificado(a). No medites mucho tu respuesta, no hay respuestas buenas ni malas. Las opciones de respuesta son:

0 = Muy en desacuerdo, 1 = En desacuerdo, 2 = Ni de acuerdo ni en desacuerdo,
3 = De acuerdo, 4 = Muy de acuerdo

	0	1	2	3	4
1. En mi opinión en la universidad no deberían enseñar investigación.					
2. En los eventos de investigación (congresos, encuentros) me relaciono con la gente.					
3. De las cosas que más me agradan son las conversaciones científicas.					
4. Eso de estar tomando cursos de actualización no es para mí.					
5. Creo que estar consultando información científica es perder el tiempo.					
6. Considero que tengo la paciencia necesaria para investigar.					
7. Todos los profesionales deberían aprender a investigar.					
8. La mayoría de las cosas me generan curiosidad.					
9. Casi siempre dejo para después lo que tiene que ver con investigación.					
10. Busco informarme de los temas de actualidad.					
11. Me gusta capacitarme para adquirir habilidades de investigación.					
12. Creo que la persistencia contribuye a alcanzar las metas.					
13. Acostumbro a escribir para profundizar en temas de interés.					
14. Las actividades de investigación diarias no me inspiran nada novedoso.					
15. Investigar es posible si tenemos interés de hacerlo.					
16. Con frecuencia me encuentro consultando información en artículos científicos.					
17. La investigación es una de las cosas que me despierta interés.					
18. Soy ordenado(a) en mis actividades de investigación.					
19. Las conversaciones científicas me parecen aburridas.					
20. Trabajar con otros en investigación nos ayuda a alcanzar mejores resultados.					
21. Se me ocurren ideas innovadoras acerca de problemas cotidianos.					
22. Considero que la investigación ayuda a detectar errores de la ciencia.					
23. Para ser sincero(a) realmente lo que menos hago es escribir.					
24. Aprovecho cualquier oportunidad para dar a conocer mis trabajos de la escuela relacionados a investigación.					
25. Me gusta agilizar los trabajos relacionados con investigación.					
26. Para mí, en investigación es importante fortalecer la capacidad de escuchar.					
27. Pensar en ponerme a investigar me produce desánimo.					
29. En mi opinión, sin investigación la ciencia no avanzaría.					
31. A mi parecer la investigación contribuye a resolver problemas sociales.					
33. Reconozco que la investigación ayuda a corregir errores del sentido común.					
34. Soy el último en enterarse de los temas de actualidad.					
35. Para ser sincero(a) realmente lo que menos hago es leer.					

Evaluación adaptativa del Neuroticismo mediante el Cuestionario Revisado de Personalidad de Eysenck

Adaptive Evaluation of Neuroticism through the Revised Eysenck Personality Questionnaire

Facundo Juan Pablo Abal^{1, 2*}, Gabriela Susana Lozzia²,
Sofía Esmeralda Auné^{1, 2}, Horacio Félix Attorresi².

1 - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

2 - Instituto de Investigaciones de la Facultad de Psicología, Universidad de Buenos Aires, Argentina.

Recibido: 13/08/2018 Revisado: 14/01/2019 Aceptado: 28/01/2019

Introducción
Objetivos
Método
Resultados
Discusión
Referencias

Resumen

Se estudió la conveniencia de administrar adaptativamente los ítems de Neuroticismo pertenecientes a la adaptación argentina del Cuestionario Revisado de Personalidad de Eysenck. Participaron 594 personas de población general. Un 70% de los casos sirvió para calibrar los ítems según el modelo logístico de dos parámetros. El resto se usó para analizar evidencias de validez y confiabilidad del test adaptativo informatizado usando un criterio de parada mixto (error de estimación $\leq .4$ o administración de 9 ítems). El 48.9% de los participantes alcanzó el error $\leq .4$ luego de presentarles 6 ítems y el 69.1% requirió de 9 ítems o menos. La correlación entre el nivel estimado con el test completo y la versión adaptativa fue de .98. Tampoco varió sustancialmente la asociación con criterios externos. La discusión aborda las ventajas y limitaciones vinculadas a la falta de ítems que discriminen en los niveles bajos del rasgo.

Palabras clave: *Neuroticismo, Eysenck, test adaptativo informatizado, modelo logístico de dos parámetros, Teoría de Respuesta al Ítem*

Abstract

This work analyses the convenience of adaptively administering the neuroticism items belonging to the Argentine adaptation of the Revised Eysenck Personality Questionnaire. The study included 594 people from the general population. Seventy percent of the cases were used to calibrate items according to the two-parameter logistic model. The rest of the cases were used to analyse evidence of validity and reliability of the computerized adaptive test using a mix stopping criterion (estimation error $\leq .4$ or administration of 9 items). Results showed that 48.9% of the participants reached an $\leq .4$ error after the administration of 6 items and that 69.1% required 9 items or less. The correlation between the level estimated with the full test and the adaptive version was of .98. The association with external criteria did not vary substantially either. The discussion addresses the advantages and limitations that stem from the lack of items which are discriminative at low-trait levels.

Key words: *Neuroticism, Eysenck, computerized adaptive test, two-parameter logistic model, Item Response Theory*

* **Correspondencia a:** Facundo Juan Pablo Abal; Dirección postal: Zuviría 5691 (1439), Ciudad Autónoma de Buenos Aires. Teléfono: 54 11 46014331. E-mail: fabal@psi.uba.ar

Cómo citar este artículo: Abal, F. J. P., Lozzia, G. S., Auné, S. E., & Attorresi, H. F. (2019). Evaluación adaptativa del Neuroticismo mediante el Cuestionario Revisado de Personalidad de Eysenck. *Revista Evaluar*, 19(1), 17-34. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Financiación: Esta investigación fue financiada con subsidios de la Universidad de Buenos Aires UBACyT 2018 Código 20020170100200BA y UBACyT 2018 Código 20020170200001BA y de la Agencia Nacional de Promoción Científica y Tecnológica PICT-2017-3226.

Introducción

El estudio conceptual e instrumental del *neuroticismo* (*N*) tiene una extensa tradición en el campo de la evaluación de la personalidad. Sin embargo, a pesar de que se registran algunas referencias más remotas (Thurstone & Thurstone, 1930; Woodworth, 1919), la intención de una medición psicométricamente válida del *N*, basada en una teoría sólida y con sustento empírico llegaría recién a partir de la década de 1940 con los desarrollos de Guilford (1940), Cattell (1957) y principalmente por la contribución de Eysenck (1947). Esto inauguró una concepción moderna del *N*, entendiéndolo como un continuo cuantitativo sobre el que es posible reconocer diferentes niveles de intensidad (Dumont, 2010; Jeronimus, Kotov, Riese, & Ormel, 2016).

El *N* fue la primera de las dimensiones identificadas por Eysenck en la construcción de su modelo de la personalidad, el cual incluye también las dimensiones *extraversión* y *psicoticismo*. La caracterización del *N* presenta en sus orígenes más lejanos la descripción de los cuatro temperamentos clásicos de la tradición hipocrática-galénica (Eysenck, 1947, 1952; Zuckerman, 2005). Siguiendo a Wundt, Eysenck vinculó el *N* con los temperamentos melancólico y colérico en tanto que para ambos se describe una tendencia a experimentar una reacción emocional de fuerte intensidad (como se citó en Bech, 2016). Los sujetos con niveles elevados de *N* son propensos a reaccionar de manera exagerada frente a todo tipo de estímulos y muestran dificultades para retornar al nivel emocional normal luego de estas experiencias. En el polo opuesto, la persona con bajo *N* es tranquila, controlada y mantiene un temperamento equilibrado en la mayoría de las circunstancias. Para Eysenck, el *N* se ubica en el nivel más alto de una estructura jerárquica y se conforma a partir de las intercorrelaciones de un

conjunto de rasgos primarios: depresión, miedo, ansiedad, preocupación, irracionalidad, tensión, timidez, vergüenza, sentimiento de culpa, baja autoestima, mal humor y emotividad (Eysenck & Eysenck, 1975). La combinación de estos aspectos en mayor o menor medida configura el nivel de estabilidad-inestabilidad emocional de la persona.

A lo largo de su obra Eysenck fue perfeccionando la medición del *N* y las otras variables de su modelo desarrollando una sucesión de cuestionarios basados en el análisis factorial (Furnham, Eysenck, & Saklofske, 2008). El Cuestionario de Personalidad Eysenck EPQ, (Eysenck & Eysenck 1975), su posterior versión revisada (EPQ-R) y su versión corta (EPQ-RS; Eysenck, Eysenck, & Barrett, 1985) podrían considerarse como el resultado final de un arduo trabajo en la construcción de una medida que operacionaliza las dimensiones básicas de personalidad que propuso y una escala de deseabilidad social (Bech, 2016). Mientras que el EPQ-R consta de 100 ítems, la versión corta emplea una selección de 48 elementos. Esta considerable reducción motivó a numerosos investigadores a efectuar adaptaciones locales de la versión corta (Aluja, García, & García, 2003; Francis, Lewis, & Ziebertz, 2006; Lewis, Francis, Shevlin, & Forrest, 2002; Squillace, Picón-Janeiro, & Schmidt, 2013; Tiwari, Singh, & Singh, 2009).

Aunque los cuestionarios de Eysenck no fueron diseñados para diagnosticar trastornos psiquiátricos, han resultado de gran utilidad en el ámbito clínico. Desde sus primeros estudios, Eysenck (1947) relacionó los niveles elevados de *N* con una predisposición a padecer lo que clásicamente se denominó trastornos neuróticos. En la actualidad el *N* es considerado como una dimensión de la personalidad normal prioritaria en virtud de su impacto en la salud pública (Vittengl,

2017; Widiger & Oltmanns, 2017). Su asociación con distintas psicopatologías ha llevado a concebirlo como un factor de vulnerabilidad común para el desarrollo y mantenimiento de diversos trastornos mentales (Jeronimus et al., 2016; Sauer-Zavala, Wilner, & Barlow, 2017). El *N* también es relevante por su vinculación con una gran variedad de enfermedades físicas, como cardiopatías, diabetes, asma o enfermedades inflamatorias intestinales (Lahey, 2009). Por esta demostrada contribución del *N* a tantas consecuencias negativas sobre la salud, en los últimos años se ha empezado a recomendar la detección de niveles altos de este rasgo en población general durante la asistencia clínica de rutina (Hengartner, Kawohl, Haker, Rössler, & Ajdacic-Gross, 2016; Tackett & Lahey, 2017; Widiger, 2009; Widiger & Oltmanns, 2017).

La demanda de una evaluación a gran escala y/o con fines de rastreo conlleva la necesidad práctica de seleccionar instrumentos cortos que permitan la detección rápida de personas que presentan un riesgo potencial (Baldasaro, Shanahan, & Bauer, 2013). Sin embargo, optimizar la medición de los rasgos no es una tarea sencilla. Son numerosos los estudios instrumentales que en los últimos años han perseguido la construcción de una medida breve para *N* así como también para otros dominios de la personalidad (Donnellan, Oswald, Baird, & Lucas, 2006; Gosling, Rentfrow, & Swann, 2003; Soto & John, 2016, 2017). Pero las estrategias para reducir la cantidad de ítems en el marco de la Teoría Clásica de Test (TCT) tienen un costo en la calidad psicométrica que no siempre es considerado (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). Frente a la validación de las escalas breves, los investigadores se enfrentan a encontrar un equilibrio razonable entre resguardar validez de contenido (Milojev, Osborne, Greaves, Barlow, & Sibley, 2013; Ziegler, Kemper, & Krueger, 2014) o precisión en la medida (Baldasaro

et al., 2013; Sibley, 2012).

Más recientemente, y al amparo del crecimiento de la Teoría de Respuesta al Ítem, se han desarrollado Test Adaptativos Informatizados (TAIs) para la medición de rasgos de la personalidad (Reise & Revicki, 2015). La implementación de una estrategia adaptativa para la medición del *N* puede resultar particularmente eficiente en la detección de niveles elevados del rasgo dado que podría ofrecer un marco de referencia psicométrico que permita acortar el cuestionario sin comprometer la precisión o validez (Chang, 2015).

Objetivos

La motivación para construir una escala breve que mida *N* no es novedosa. En efecto, Eysenck (1958) elaboró tempranamente una primera escala con apenas seis ítems. Posteriormente, algunos autores se propusieron acortar el EPQ-RS desde la TCT y ensayaron versiones abreviadas (EPQ-RA) que también utilizan seis ítems (Francis, Brown, & Philipchalk, 1992; Sandín, Valiente, Chorot, Olmedo, & Santed, 2002). Pero la reducción en la cantidad de reactivos tuvo como consecuencia una irremediable disminución en la confiabilidad de la escala (Lewis et al., 2002).

La medición de *N* mediante EPQ-RS acumula sólidas propiedades psicométricas tanto a nivel local como internacional. Considerando las ventajas que brindaría la evaluación adaptativa de este constructo en un ámbito clínico-epidemiológico y a la luz de los avances metodológicos mencionados, se plantean los siguientes objetivos: a) evaluar si los 12 ítems de *N* de la adaptación argentina de la EPQ-RS cumplen con los requisitos psicométricos necesarios para ser administrados adaptativamente y b) examinar si la versión adaptativa de esta escala podría mostrar ventajas por sobre la administración

tradicional.

Método

Participantes

Se contó con la colaboración de 594 personas residentes en el área metropolitana de Buenos Aires, Argentina. El 52% pertenecían al género femenino y la edad osciló entre 18 y 82 años alcanzando un valor promedio de 32.9 años (DE = 12.1). El 79% consignó estar empleado y el 31% se encontraba cursando una carrera terciaria o universitaria al momento de ser evaluado. El 54.1% reportó tener un nivel de estudios secundarios completo y sólo un 6.2% no alcanzó este nivel. Un 18.5% presentó estudios terciarios (completos o incompletos) y un 21.2% registró un nivel universitario (completo o incompleto).

Instrumentos

Escala Neuroticismo del Eysenck Personality Questionnaire revised short version, EPQ-RS (Eysenck et al., 1985; adaptación de Squillace et al., 2013). El cuestionario original consta de 48 ítems, pero en la adaptación argentina quedó reducido a 42. De estos, sólo 12 son utilizados para medir el *N*. Los ítems están redactados en forma de pregunta y el formato de respuesta es dicotómico (*Sí-No*). Los adaptadores del instrumento replicaron la estructura factorial propuesta por Eysenck y registraron un índice de consistencia interna satisfactorio (KR-20 = .84) para la escala *N*. En el presente estudio se halló un índice, ligeramente más alto, de .85 para la muestra de calibración.

Inventario de síntomas SCL-90-R (Derogatis, 1994). Es un autoinforme compuesto por 90 ítems que evalúan patrones de síntomas psicológicos con el fin de detectar personas en

riesgo. Los elementos presentan un formato de respuesta politómica en la que el sujeto responde si el síntoma le ha preocupado o molestado *nada, muy poco, poco, bastante* o *mucho* durante los últimos siete días. Los ítems se agrupan para posibilitar la medición de nueve dimensiones clínicas (*somatización, obsesiones/compulsiones, sensibilidad interpersonal, depresión, ansiedad, hostilidad, ansiedad fóbica, ideación paranoide y psicoticismo*) y tres índices globales (*índice de severidad global, total de síntomas positivos e índice de malestar positivo*). Los estudios de la adaptación argentina de este inventario muestran suficientes evidencias de validez y confiabilidad tanto en población clínica (Sánchez & Ledesma, 2009) y no clínica (Casullo, 2004). En cuanto al presente estudio, el análisis de la consistencia interna de todos los ítems del inventario mostró un coeficiente alfa de Cronbach de .96 mientras que para las dimensiones clínicas este coeficiente osciló entre .77 y .86.

Procedimiento

Los individuos respondieron el protocolo en formato lápiz y papel sin tiempo límite. Antes de comenzar la administración se les explicó que la prueba tenía como objetivo la medición de aspectos de su personalidad y que no había respuestas correctas o incorrectas. Se les comunicó sobre las garantías de anonimato y confidencialidad de sus respuestas. También se les informó sobre el carácter voluntario de su participación y la posibilidad de abandonar la evaluación en cualquier momento de la actividad. Todas estas consideraciones fueron precisadas por escrito y formaron parte de la redacción del consentimiento que debieron firmar los sujetos para participar. El protocolo administrado incluyó otras escalas no consideradas para los fines de esta investigación.

Análisis de datos

Se dividió a los participantes en dos grupos de manera aleatoria. Las respuestas de un 70% de los sujetos ($n = 416$) fueron empleadas para calibrar los ítems según el modelo logístico de dos parámetros (ML2p) mientras que el resto de los individuos ($n = 178$) fueron considerados exclusivamente para analizar el TAI.

Calibración de los ítems

Supuestos del ML2p. Se realizó un análisis factorial confirmatorio (AFC) con el programa Mplus (Muthén & Muthén, 2010) a fin de verificar el supuesto de unidimensionalidad. Se estimaron los parámetros con el método robusto de mínimos cuadrados ponderados sobre la base de la matriz de correlaciones tetracóricas. Para analizar el ajuste se aplicaron los criterios recomendados por Byrne (2012) para los índices de ajuste comparativo ($CFI \geq .90$) y de Tucker-Lewis ($TLI \geq .90$) y el error medio cuadrático de aproximación ($RMSEA \leq .08$).

Si bien es habitual dar por satisfecho el supuesto de independencia local de los ítems a partir de verificar la unidimensionalidad del constructo, los tests de personalidad construidos originalmente desde la perspectiva de la teoría clásica suelen incluir ítems redundantes que ocasionan problemas de dependencia local (Reise & Revicki, 2015). Es por esta razón que, habiendo procedimientos específicos para su comprobación, se decidió estudiar el supuesto de independencia local. Su análisis se implementó mediante los estadísticos χ^2 que suministra el programa MODFIT (Stark, 2001) para contrastar la independencia a partir de pares y tríos de ítems. Para hacer manejable el número de combinatorias de pares

$\binom{n}{2}\binom{n}{2}$ y tríos $\binom{n}{3}\binom{n}{3}$ de ítems, el programa analizó la dependencia dividiendo a los 12

ítems de la prueba en cuatro conjuntos de tres ítems (Drasgow, Levine, Tsien, Williams, & Mead, 1995). Posteriormente, se calcularon los estadísticos χ^2 para cada ítem (simple), para todas las combinaciones de pares posibles al interior de cada conjunto de ítems (12 pares) y para los cuatros tríos. Los valores del ratio χ^2/gl superiores a 3 reflejan problemas de desajuste del modelo a los datos ocasionados por la dependencia local entre los ítems (Abad, Olea, Ponsoda, & García, 2011).

Estimación y Ajuste del ML2p. El ML2p supone que la probabilidad $P_i(\theta)$ de que un individuo con un nivel de rasgo θ brinde una determinada respuesta a un ítem dicotómico está en función de los parámetros de localización b_i y de discriminación a_i del ítem. Su formulación viene dada por la expresión:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7 a_i(\theta - b_i)}}$$

La aplicación del ML2p se implementó con MULTILOG (Thissen, 2003). La estimación de los parámetros se efectuó por máxima verosimilitud marginal (MML) con un criterio de convergencia de .0001. Se estimaron 12 parámetros de localización y 12 parámetros de discriminación. Así también se estimaron valores de θ para cada sujeto sobre la base de sus patrones de respuesta a la prueba completa. Se examinó el funcionamiento diferencial de los ítems (DIF) en función del género a partir del test de Wald modificado, realizado con el software IRTPRO (Cai, Thissen, & Du Toit, 2011).

El ajuste se analizó a partir de los métodos gráficos que brinda MODFIT. El programa proporciona una comparación de las probabilidades observadas y esperadas para 25 niveles del rasgo permitiendo establecer si el modelo predice adecuadamente las curvas

empíricas. Adicionalmente, se estudiaron otros indicadores indirectos de ajuste recomendados por Rubio, Aguado, Hontangas y Hernández (2007): a) una convergencia en la estimación con una cantidad razonable de iteraciones, b) valores estimados acordes a lo esperable, c) errores estándar relativamente bajos.

Determinación del algoritmo adaptativo.

Los ítems calibrados fueron usados como insumos para el estudio del TAI. Se efectuó una simulación sobre datos reales (simulación post hoc) operando el programa Firestar-D (Choi, 2009). A partir de la matriz de datos recabados el software selecciona progresivamente aquellas respuestas correspondientes a los ítems que presentaría en el caso de estar respondiendo un TAI.

Se consideró la media del rasgo como una estimación inicial razonable de θ para dar comienzo a la administración (Embretson & Reise, 2000). Luego de la respuesta a cada ítem, la estimación provisional de θ se realizó con el método bayesiano de estimación esperada a posteriori (EAP) utilizando a la normal estándar como distribución a priori. En la selección progresiva de los ítems se aplicó el método de máxima información de Fisher consistente en presentar al individuo el ítem que resulte más informativo para el nivel provisional de θ estimado hasta ese momento. En este punto se añadió la especificación de que la selección se realizó eligiendo al azar uno de los tres ítems con máxima función de información. Se esperó con esto alcanzar una mejor representatividad del contenido en el muestreo de los ítems. Se definió un criterio de parada mixto por el cual la administración de ítems finalizaría al conseguirse un nivel de error en la estimación del rasgo igual o inferior a .4 o al presentarse 9 ítems (25% menos de ítems que los utilizados en la versión

convencional del test).

Para determinar si este algoritmo adaptativo brinda un nivel de rasgo aproximado al que obtendría la persona al responder a todos los ítems, se correlacionaron los niveles θ estimados a partir de este TAI con los estimados al responder todos los ítems. Además, se analizó si este tipo de administración podría afectar las evidencias de validez convergente con las escalas del SCL-90-R. Por ello, se aplicaron pruebas de diferencia de r de Pearson para muestras relacionadas (Steiger, 1980) con el fin de comparar si los coeficientes de correlación con estas escalas diferían al estimar las puntuaciones de N con el TAI o con el test completo.

Resultados

Fase de calibración de los ítems

Supuestos del ML2p. Los índices de ajuste comparativos evidenciaron valores adecuados (TLI = .94 y CFI = .95). Si se considera el índice de ajuste absoluto RMSEA = .075 (IC 90% .063 - .087), el límite superior del IC podría suponer cierto desajuste. No obstante, en términos generales, es posible aceptar razonablemente la unidimensionalidad del constructo a partir de estos resultados. Todas las cargas factoriales para la solución unidimensional fueron estadísticamente significativas ($p < .05$) con valores que variaron entre .51 (ítem 3) y .89 (ítem 11). Con respecto al supuesto de independencia local, en la Tabla 1 se reproduce la distribución de frecuencias y resúmenes estadísticos de los valores de las ratio χ^2/gl obtenidos con MODFIT. Como se puede observar, la contrastación de la independencia local de los ítems por pares y tríos mostró que todas las ratios resultaron menores a 3, apuntando a la verificación de este supuesto.

Tabla 1

Frecuencias y estadísticos descriptivos de la ratio χ^2/gl para contrastar la independencia local de los ítems.

	Tabla de frecuencias			Media	DE
	< 1	1 < 2	2 < 3		
Simple	12	0	0	0.003	0.006
Pares de ítems	8	4	0	0.622	0.523
Tríos de ítems	1	3	0	1.066	0.612

Aplicación del ML2p. El proceso de estimación requirió de 26 iteraciones para alcanzar el criterio de convergencia de .0001, lo cual resulta en una cantidad de ciclos razonable. Los parámetros estimados para los ítems presentaron valores dentro de un rango esperable y los errores de estimación resultaron relativamente bajos (Tabla 2).

El valor promedio de los parámetros de localización de los 12 ítems fue de 0.24 (DE = 0.27) y sus valores mínimo y máximo fueron de -0.04 (*harto*) y 0.88 (*sufriente*). Estos resultados indican que los ítems se localizaron con una fuerte concentración en los niveles medio-altos del rasgo. Se constató una significativa falta de ítems que permitan registrar las diferencias individuales de personas con niveles bajos del rasgo. A nivel descriptivo es posible señalar que los contenidos de los ítems demandan un nivel de N relativamente elevado como para tener más chances de responder en el sentido del rasgo. En efecto, la distribución de los θ estimados para los sujetos osciló entre -1.31 y 0.72 con una media de -0.27 (DE = 0.63) lo que denota una tendencia de los evaluados a puntuar en niveles medio-bajos de N .

Tabla 2

Parámetros de los ítems, estudio de DIF y uso en el Test Adaptativo Informatizado (TAI).

Ítems	Parámetros		Test de Wald			% de uso en el TAI
	b (<i>se</i>)	a (<i>se</i>)	χ^2	gl	p	
1) Altibajos en ánimo	-0.02 (.07)	1.58 (.32)	0.9	2	.6232	82
2) Desdichado sin motivo	0.21 (.09)	1.15 (.25)	0.9	2	.6332	83
3) Irritable	0.33 (.15)	0.58 (.17)	0.2	2	.9166	20
4) Harto	-0.04 (.09)	1.19 (.26)	2.2	2	.3276	74
5) Culpa	0.11 (.10)	0.94 (.22)	0.3	2	.8682	61
6) Nervioso	0.14 (.12)	0.73 (.19)	0.3	2	.8619	27
7) Sufriente	0.88 (.08)	1.82 (.48)	2.0	2	.3626	45
8) Tensa	0.59 (.12)	0.90 (.23)	1.3	2	.5113	35
9) Desgano	0.09 (.08)	1.22 (.26)	3.9	2	.1432	90
10) Vida monótona	-0.04 (.11)	0.80 (.19)	0.2	2	.9015	35
11) Inestable	0.28 (.07)	1.72 (.37)	1.8	2	.4039	89
12) Sentimiento de soledad	0.33 (.09)	1.17 (.27)	0.5	2	.7739	79

Nota. Ítem: Expresión que sintetiza el indicador indagado por cada ítem; *se*: Error de estimación.

Los parámetros de discriminación mostraron un promedio de 1.15 (DE=0.39), lo que supone que los ítems presentan una potencia discriminativa

moderada en torno a sus respectivos valores de b . El único ítem que presentó un valor bajo de discriminación indaga sobre la tendencia del individuo a enojarse fácilmente ($a = .58$). Entre

los ítems con mayores valores de a se destacan aquellos que preguntan por cambios en el estado del ánimo y por percibirse como una persona que sufre mucho.

En la Tabla 2 también se aporta información sobre el estudio del DIF de los ítems en función del género. Los valores p asociados al estadístico χ^2 del test de Wald señalan que no se observan diferencias estadísticamente significativas en los

parámetros de los ítems estimados para mujeres y varones.

En la Figura 1 se muestran los gráficos de ajuste que proporciona MODFIT para cada ítem. Para todos los elementos se verifica que la curva característica del ítem se encuentre dentro del intervalo de confianza asociado a la probabilidad observada de los distintos niveles del rasgo que son contrastados.

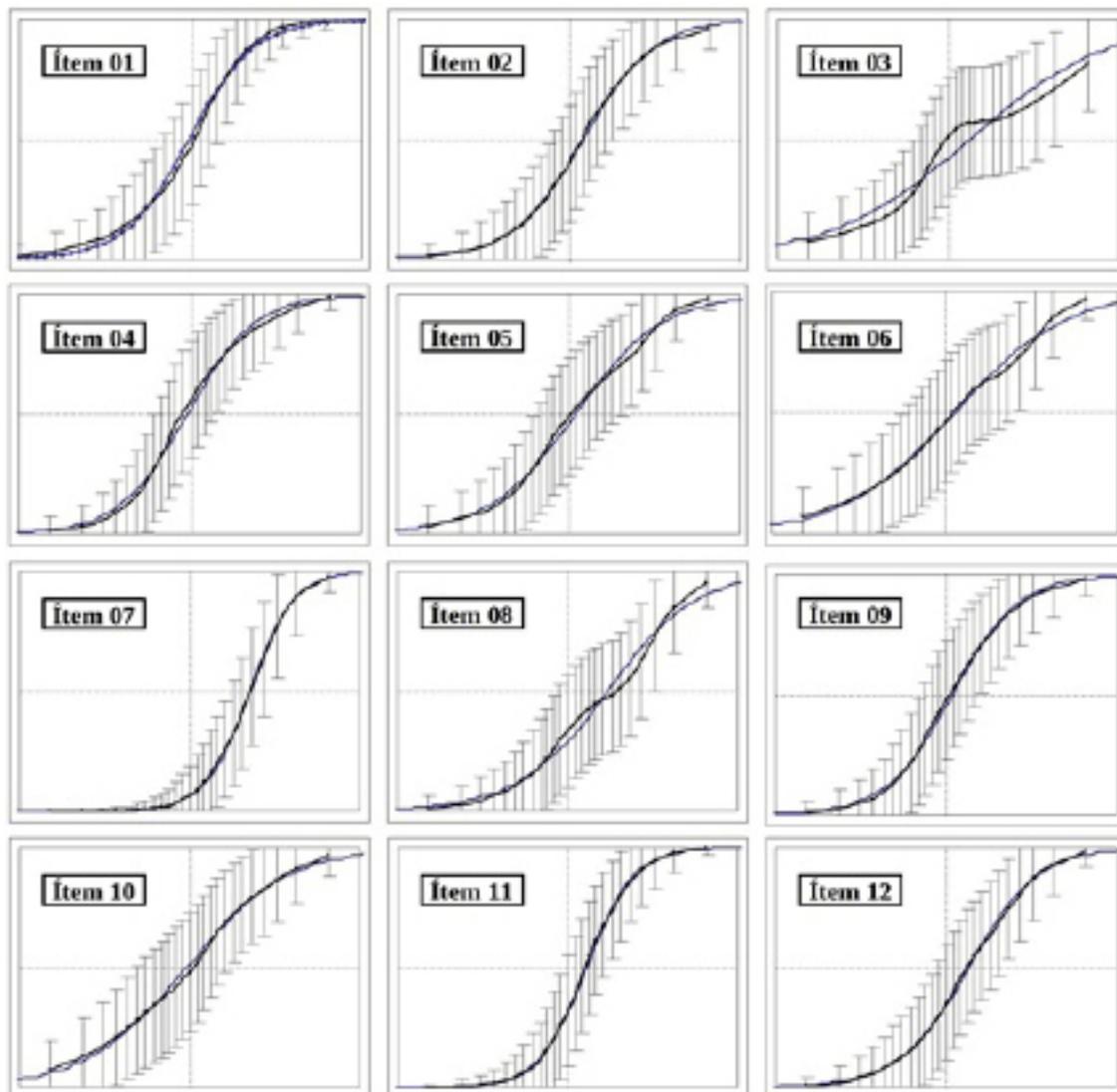


Figura 1

Gráficos de ajuste para los ítems de la Escala de Neuroticismo al Modelo Logístico de dos Parámetros.

Fase de simulación del TAI

Al analizar los resultados obtenidos tras la simulación del TAI con criterio de parada mixto se encontró una correlación alta y positiva entre los valores θ estimados con el test completo y su versión adaptativa ($r = .98$; $p < .001$) y también entre los errores de estimación respectivos ($r = .77$; $p < .001$). La asociación resultó también muy elevada al correlacionar el TAI y el puntaje total obtenido desde la perspectiva de la clásica ($r = .96$; $p < .001$). La diferencia media entre los niveles de rasgo estimados para los evaluados en el TAI y sus correspondientes niveles estimados a partir del test completo (sesgo empírico) resultó de .018.

En las condiciones que establece el criterio de parada adoptado se requirió administrar un promedio de 7.2 ítems por sujeto. El 48.9% de los participantes alcanzó un error de $\leq .4$ luego de presentarles 6 ítems y el 69.1% requirieron de 9 ítems o menos. El 30.9% restante ($n = 55$)

no alcanzó el error prefijado y se interrumpió su evaluación por alcanzar el tope de cantidad de ítems.

Los errores de estimación variaron entre .35 y .47 con una media de .39 (DE = .032). Esto implica que se alcanzó un nivel de precisión adecuado considerando la reducción en la cantidad de ítems administrados. El error máximo (.47) en la estimación del θ del TAI se corresponde con una confiabilidad clásica de .78. Esto significa que ningún individuo fue evaluado con confiabilidad menor que .78. El error promedio es equivalente a una confiabilidad de .85. Esto supone que para la mayoría de los sujetos evaluados la confiabilidad del TAI, en términos clásicos, fue similar a la que se obtuvo con el test en su formato convencional, pero requirió administrar entre un 50% y 75% menos de ítems. Asimismo, los casos en los que la estimación se realizó con mayor error no son mejor evaluables con la versión completa de la escala ya que no hay ítems que aporten máxima información en esos niveles, como se verá a continuación.

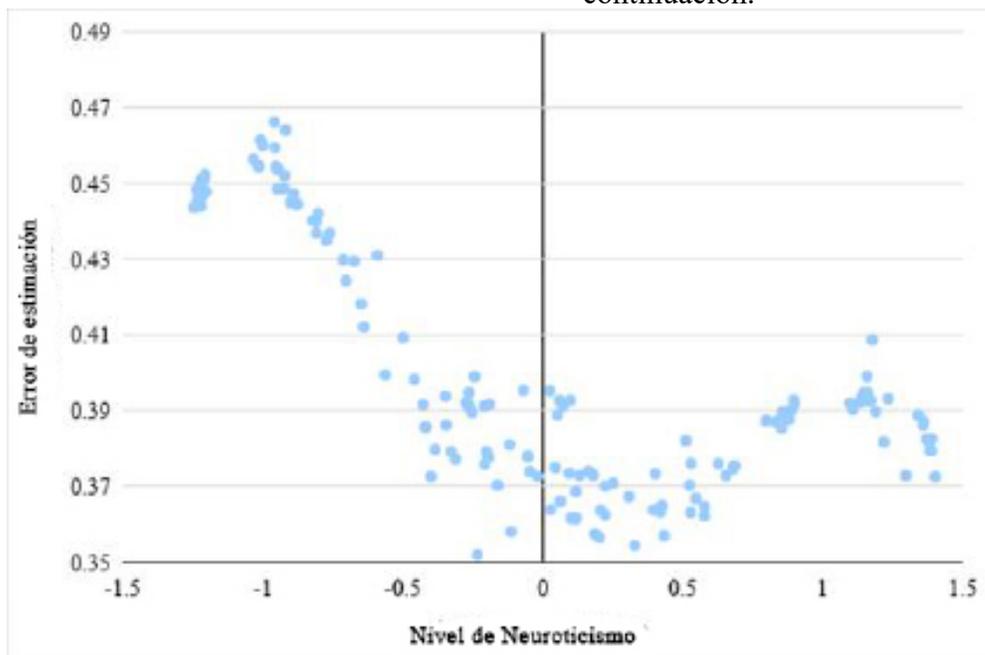


Figura 2

Diagrama de dispersión de los errores de estimación en función de los niveles de rasgo estimados a partir del TAI.

Se obtuvo una correlación de $-.72$ ($p < .001$) entre el nivel de rasgo estimado y el error, indicando que a menor θ estimado se verificó un mayor error de estimación. Este hallazgo era esperable ya que los ítems de la prueba se localizan en un rango restringido del rasgo por encima de la media de θ . La Figura 2 exhibe el diagrama de dispersión de los errores de estimación en función del θ estimado con el TAI. Allí se observa que la mayoría de los sujetos con niveles de $\theta \geq -.5$ presentan un error de $\leq .4$. En cambio, los sujetos que no alcanzaron el error de estimación prefijado al finalizar la administración adaptativa mostraron un θ promedio de $-.96$ (DE = $.36$) y un error promedio de $.44$ (DE = $.0012$).

Los evaluados con menores niveles de rasgo también fueron los que necesitaron mayor cantidad de ítems y, sin embargo, no alcanzaron el nivel de precisión esperado. En este sentido, la finalización de la administración en 9 ítems parece justificada para efectuar un uso más eficiente del tiempo y del esfuerzo del evaluado. Los tres ítems adicionales que le permitirían al sujeto completar el instrumento aportaron muy poca información para la medición dado que se encontraban alejados de la localización del individuo en la escala del rasgo. En la Tabla 2 se observa el porcentaje de casos en los que fue administrado cada ítem. Los ítems menos usados son los que presentaron menor capacidad discriminativa como el 3 (*irritable*) y el 6 (*nerviosismo*). Entre los más aplicados aparecen aquellos que tienen parámetros a entre moderado y alto y cuyos parámetros b de localización fueron próximos a la media del rasgo. El ítem 7 (*sufriente*) presentó el valor más elevado de discriminación, pero solo fue usado en el 35% de los casos porque se localiza en un nivel medio-alto de la escala. Debido a que el TAI se iniciaba con un ítem seleccionado al azar entre los tres más informativos para el nivel de rasgo

medio, los ítems 9 (*desgano*), 11 (*inestable*) y 1 (*altibajos en el ánimo*) fueron los utilizados para arrancar la administración.

Finalmente, el estudio de la asociación entre el N y las escalas del SCL-90-R mostró resultados acordes a lo esperable desde una perspectiva teórica (Tabla 3). Aun remitiendo a la personalidad normal, el N es un rasgo que describe una predisposición a padecer trastornos psicopatológicos. En este sentido, es coherente que los individuos tiendan a manifestar mayor intensidad y/o variedad de sintomatología psicológica cuanto mayor nivel de N experimenten. Si bien los resultados van en este sentido, el hallazgo más destacable que se observa en la Tabla 3 es que las evidencias de validez convergente con las escalas del SCL-90-R no registraron diferencias estadísticamente significativas al ser estudiadas con las estimaciones de θ surgidas del TAI o del test completo. Esto significa que la reducción en la cantidad de los ítems que presenta la versión adaptativa no impactó sustantivamente en estas evidencias de validez.

Discusión

Si bien pareciera que el modelo de personalidad de Eysenck ha sido opacado por la preponderancia actual del modelo de los Cinco Factores (Digman, 1990; Goldberg, 1999; McCrae & Costa, 2010), el uso de sus cuestionarios no ha perdido vigencia ni en su versión para población adulta (Almiro, Moura, & Simões, 2016; Colledani, Anselmi, & Robusto, 2018; Moeller et al., 2015; Picconi, Jackson, Balsamo, Tommasi, & Saggino, 2018; Zambrano-Cruz, 2011) ni en su versión para adolescentes (Alcázar-Córcoles, Verdejo-García, Bouso-Sáiz,

Tabla 3

Correlación de las estimaciones del TAI y el test completo con criterios externos.

Escalas del SCL90-R	TAI	Test completo	Prueba de diferencia de r de Pearson
Somatizaciones	.40**	.40**	$t(175) = .36$; ns
Obsesiones y Compulsiones	.51**	.52**	$t(175) = .54$; ns
Sensitividad Interpersonal	.50**	.51**	$t(175) = .46$; ns
Depresión	.65**	.66**	$t(175) = .56$; ns
Ansiedad	.56**	.57**	$t(175) = .64$; ns
Hostilidad	.43**	.43**	$t(175) = .37$; ns
Ansiedad Fóbica	.38**	.38**	$t(175) = .21$; ns
Ideación Paranoide	.36**	.36**	$t(175) = .28$; ns
Psicoticismo	.45**	.45**	$t(175) = .45$; ns
Índice de Severidad Global	.52**	.53**	$t(175) = .55$; ns
Total de Síntomas Positivos	.49**	.50**	$t(175) = .53$; ns
Índice de Malestar Positivo	.52**	.53**	$t(175) = .55$; ns

Nota. ** $p < .001$.

Reuelta-Menéndez, & Ramírez-Lira, 2017; Colledani, Robusto, & Anselmi, 2018). Esto se debe fundamentalmente a que da como resultado un modelo paradigmático causal de corte biologicista que cuenta con un importante volumen de evidencia experimental acumulada (Flores-Mendoza et al., 2016; Revelle, 2016). Incluso los estudios recientes sobre una fundamentación genética del N (Calboli et al., 2010; Luciano et al., 2018; Mitchell & Kumari, 2016; Smith et al., 2016) y la replicación del modelo en diversas culturas (Bowden, Saklofske, van de Vijver, Sudarshan, & Eysenck, 2016; Eysenck & Barrett, 2013) muestran cómo se sostiene la preocupación de Eysenck por encontrar bases empíricas que complementen los hallazgos obtenidos con el análisis factorial de los cuestionarios.

Dentro del modelo psicobiológico de Eysenck, la dimensión N presenta atributos relevantes que justifican su selección para los fines de este trabajo. En el plano teórico, N se

destaca por haber sido caracterizado con relativa concordancia por las diversas teorías factorialistas de la personalidad contemporáneas a Eysenck y por ser reconocido por desarrollos posteriores (De Raad & Mlačić, 2015; Zuckerman, 2005). Existen pocos modelos empíricos o teóricos de la estructura general de la personalidad que no incluyan un dominio como N (Digman, 1990). A su vez, en el plano aplicado, la vinculación de N con distintos trastornos psicopatológicos y enfermedades físicas ofrece como ventaja la posibilidad de pensar estrategias transdiagnósticas de detección y prevención (Hajek, Bock, & König, 2017; Lahey, 2009; Vittengl, 2017; Widiger & Oltmanns, 2017).

La metodología TAI aquí implementada ha demostrado proporcionar estimaciones de los niveles de N de los evaluados con un grado de precisión aceptable aun cuando se administra sólo una parte de los ítems que componen la escala N del EPQ-RS. Similares resultados se

han registrado al estudiar evidencias de validez de versiones adaptativas de tests reconocidos como el MMPI-2 (Forbey & Ben-Porath, 2007; Salazar-Schneiderman, Tonetti, Sommers, & Greene, 2016) y el NEO-PI-R (Reise & Henson, 2000) y se encontraron idénticos estándares de calidad de la medida administrando solo la mitad de reactivos que la equivalente convencional. La disminución de la cantidad de ítems de N del EPQ-RS administrados puede variar entre un 50% y 75% dependiendo de la localización del sujeto en la dimensión del rasgo.

Un aspecto que conviene discutir es la posibilidad de que, al administrar una menor cantidad de ítems con el TAI, se vea afectada la representatividad del contenido muestreado. La validez de contenido del TAI está supeditada a los ítems con los que cuenta el EPQ-RS. El algoritmo adaptativo selecciona el ítem de entre todos los disponibles basándose en un criterio de calidad psicométrica. Su administración dependerá de la información que pueda aportar el ítem por su cercanía al nivel del rasgo estimado para ese sujeto independientemente del contenido al que alude el indicador. Con respecto a la implementación adaptativa del EPQ-RS, la correlación elevada observada entre el θ estimado con el TAI y el θ obtenido con el test completo demuestra que las variaciones en el contenido no afectaron sustantivamente la medición del constructo. Tampoco se alteró la relación del N con las dimensiones sintomáticas evaluadas mediante la SCL-90-R. En consecuencia, la infra o supra representación de un rasgo primario durante la medición adaptativa de una persona no tuvo consecuencias en la estimación de su rasgo.

Si se quisiera alcanzar un mayor control sobre el contenido de los ítems deberían identificarse con más precisión los rasgos primarios del N e incorporarse más elementos para la medición

de estos aspectos. Estudios posteriores a la construcción del EPQ-R llevaron a Eysenck y sus colaboradores a definir siete rasgos primarios: inferioridad, infelicidad, ansiedad, dependencia, hipocondría, culpa y obsesión (Eysenck, Barrett, Wilson, & Jackson, 1992), pero luego se consideró que los primeros tres eran los más descriptivos de la dimensión (Eysenck, Wilson, & Jackson, 1996). Obsesión se encontró débilmente asociado al resto de los rasgos mientras que dependencia, hipocondría y culpa ofrecían información redundante (Jackson, Furnham, Forde, & Cotter, 2000). La delimitación conceptual precisa del N y la identificación de sus rasgos primarios son objeto de debate actual al interior de las teorías factorialistas de la personalidad (Tackett & Lahey, 2017; Widiger, 2009).

La concentración de los parámetros de localización de los ítems en los niveles medio-altos del rasgo latente podría revelar una fortaleza o una debilidad de la escala, pero esto dependerá de los objetivos para los que se la pretenda usar. Si la finalidad de la evaluación es estudiar la distribución del rasgo en personalidad normal se hallarán dificultades en la especificación del extremo inferior en la bipolaridad del constructo. A nivel conceptual este polo se define como la tendencia a permanecer estable y tranquilo incluso al ser confrontado por estresores importantes (Furnham et al., 2008). La falta de ítems localizados en los niveles bajos de la escala imposibilita describir los rasgos distintivos que caracterizan ese rango de la variable. Esto significa que a nivel operacional la especificación de la estabilidad emocional se circunscribe a la ausencia de manifestaciones relacionadas con un nivel elevado de N . Como consecuencia, la estimación del nivel del rasgo de los sujetos próximos al polo de la estabilidad emocional resulta más imprecisa. Esta limitación es un problema de la escala de partida y no es una

consecuencia del TAI.

Los ítems del instrumento resultan más efectivos para detectar niveles elevados de N , que son justamente las puntuaciones que presentan más relevancia en el contexto clínico-epidemiológico. Esto significa que la localización de los ítems en niveles medio-altos del rasgo beneficia especialmente a la aplicación de la prueba para tareas de evaluación con fines de rastillaje. En esta línea, la implementación mediante un formato de administración adaptativo contribuye a realizar esta evaluación de manera más eficiente en dos sentidos. En primer lugar, la reducción en la cantidad de ítems posibilita que el tiempo ahorrado pueda ser aprovechado para la medición de otras variables o una simplificación de la administración que evite la fatiga innecesaria del evaluado. El ahorro de tiempo podría ser considerado despreciable en la práctica cotidiana de una evaluación individual o en grupos reducidos. Sin embargo, la economía de los tiempos de administración se vuelve significativa si la finalidad es evaluar a un conjunto numeroso de personas o en estudios a gran escala (Baldasaro et al., 2013).

En segundo lugar, el procedimiento adaptativo es más eficiente en comparación con una prueba de igual cantidad de ítems pero con formato fijo. La evaluación adaptativa selecciona los ítems para ofrecer garantías de una medición más precisa incluso en los valores extremos del rasgo. Estos niveles del rasgo son los que presentan mayor error si se emplean las pruebas convencionales (Olea & Ponsoda, 2013; Reise & Revicki, 2015; Rubio et al., 2007).

El algoritmo adaptativo ensayado en el presente estudio propone una manera de reducir la cantidad de ítems de la EPQ-RS desde una aproximación alternativa a la aplicada en estudios que se basan en la Teoría Clásica de Tests. La

selección de los ítems que componen los tests con formato fijo se realiza en función de identificar los elementos con mayor capacidad discriminativa general (Francis et al., 1992; Sandín et al., 2002). Sin embargo, en el marco de la TRI, la discriminación de un ítem se considera en torno al valor de su parámetro b . El ítem que se localiza en un nivel del rasgo en el que se hayan posicionado pocos sujetos será escasamente utilizado en el TAI a pesar de tener una elevada capacidad discriminativa (como sucedió con el ítem 7). También podría considerarse una diferencia con respecto a la validez de contenido. En las versiones breves con un formato fijo se acota el muestreo de indicadores evaluados, pero se corre el peligro de que conlleve a restringir posibles ramificaciones del constructo. En cambio, en la versión adaptativa existe la posibilidad de que cualquiera de los ítems de la prueba original sea administrado siempre que el nivel del individuo se halle cerca de su localización.

Aun cuando los resultados son alentadores es fundamental rescatar el carácter preliminar de estos. El escaso número de elementos que componen la escala limitan considerablemente la posibilidad de alcanzar una medida más precisa y eficiente. Es habitual que las mediciones adaptativas se realicen a partir de un banco que incluya una mayor cantidad de ítems (Reise & Revicki, 2015). En el presente estudio se ha recurrido a un conjunto de ítems pertenecientes a un instrumento corto pero con fuertes evidencias de validez y estudios de adaptación local. Se ha demostrado que a pesar de su escaso número de ítems es posible contar con una versión adaptativa de esta escala.

En futuros estudios se buscará incorporar nuevos ítems tendientes a identificar indicadores del constructo que detecten niveles bajos de N . Esto no resulta una tarea sencilla. Eysenck y Eysenck

(1969) concluyeron que es desaconsejable emplear ítems redactados en sentido negativo al rasgo para medir el *N*. Además, el estudio de la unidimensionalidad requiere la inversión en la codificación de las respuestas de los ítems negativos, por lo que no es seguro que queden localizados en los niveles bajos del rasgo. Parece recomendable apelar a ítems cuyos indicadores sean plasmados en los enunciados con una frecuencia de aparición más esporádica o con una manifestación de menor intensidad. Es complejo poder predecir en qué nivel de rasgo se localizará el ítem, pero se debería recurrir a matices más sutiles en la redacción de los enunciados.

Referencias

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid, España: Síntesis.
- Alcázar-Córcoles, M. A., Verdejo-García, A. J., Bouso-Sáiz, J. C., Revuelta-Menéndez, J., & Ramírez-Lira, E. (2017). Propiedades psicométricas del cuestionario de personalidad EPQ-A en una muestra de adolescentes hispanohablantes. *Anuario de Psicología Jurídica*, 27(1), 51-56. doi: [10.1016/j.apj.2017.02.002](https://doi.org/10.1016/j.apj.2017.02.002)
- Almiro, P. A., Moura, O., & Simões, M. R. (2016). Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire - Revised (EPQ-R). *Personality and Individual Differences*, 88, 88-93. doi: [10.1016/j.paid.2015.08.050](https://doi.org/10.1016/j.paid.2015.08.050)
- Aluja, A., García, O., & García, L. F. (2003). A psychometric analysis of the revised Eysenck Personality Questionnaire short scale. *Personality and Individual Differences*, 35(2), 449-460. doi: [10.1016/S0191-8869\(02\)00206-4](https://doi.org/10.1016/S0191-8869(02)00206-4)
- Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the Mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment*, 95(1), 74-84. doi: [10.1080/00223891.2012.700466](https://doi.org/10.1080/00223891.2012.700466)
- Bech, P. (2016). Neuroticism (Eysenck's Theory). En V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1-4). doi: [10.1007/978-3-319-28099-8_1094-1](https://doi.org/10.1007/978-3-319-28099-8_1094-1)
- Bowden, S. C., Saklofske, D. H., van de Vijver, F. J. R., Sudarshan, N. J., & Eysenck, S. B. G. (2016). Cross-cultural measurement invariance of the Eysenck Personality Questionnaire across 33 countries. *Personality and Individual Differences*, 103, 53-60. doi: [10.1016/j.paid.2016.04.028](https://doi.org/10.1016/j.paid.2016.04.028)
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basics, concepts, applications, and programming*. New York, NY: Routledge. doi: [10.4324/9780203807644](https://doi.org/10.4324/9780203807644)
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [software de cómputo]. Lincolnwood, IL: Scientific Software International.
- Calboli, F. C. F., Tozzi, F., Galwey, N. W., Antoniadis, A., Mooser, V., Preisig, M., ... Balding, D. J. (2010). A genome-wide association study of neuroticism in a population-based sample. *PLoS One*, 5(7), 1-7. doi: [10.1371/journal.pone.0011504](https://doi.org/10.1371/journal.pone.0011504)
- Casullo, M. (2004). Síntomas psicopatológicos en adultos urbanos. *Psicología y Ciencia Social*, 6(1), 49-57. Recuperado de <https://biblat.unam.mx/es/revista/psicologia-y-ciencia-social>
- Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*. Oxford, Inglaterra: World Book Co.
- Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80(1), 1-20. doi: [10.1007/s11336-014-9401-5](https://doi.org/10.1007/s11336-014-9401-5)
- Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for Polytomous Item Response Theory models. *Applied Psychological Measurement*, 33(8), 644-645. doi: [10.1177/0146621608329892](https://doi.org/10.1177/0146621608329892)
- Colledani, D., Anselmi, P., & Robusto, E. (2018). Using Item Response Theory for the development of a new

- short form of the Eysenck Personality Questionnaire-Revised. *Frontiers in Psychology*, 9, 1-13. doi: [10.3389/fpsyg.2018.01834](https://doi.org/10.3389/fpsyg.2018.01834)
- Colledani, D., Robusto, E., & Anselmi, P. (2018). Development of a new abbreviated form of the Junior Eysenck Personality Questionnaire-Revised. *Personality and Individual Differences*, 120, 159-165. doi: [10.1016/j.paid.2017.08.037](https://doi.org/10.1016/j.paid.2017.08.037)
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, 102(4), 874-888. doi: [10.1037/a0027403](https://doi.org/10.1037/a0027403)
- De Raad, B., & Mlačić, B. (2015). Big Five Factor model, theory and structure. En J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2^a. ed., Vol. II, pp. 559-566). Oxford, Reino Unido: Elsevier. doi: [10.1016/b978-0-08-097086-8.25066-6](https://doi.org/10.1016/b978-0-08-097086-8.25066-6)
- Derogatis, L. (1994). *SCL-90-R. Symptom Checklist-90-R. Administration, Scoring and Procedures Manual*. Minneapolis, MN: National Computer System.
- Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor model. *Annual Review of Psychology*, 41(1), 417-440. doi: [10.1146/annurev.psych.41.1.417](https://doi.org/10.1146/annurev.psych.41.1.417)
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five Factors of personality. *Psychological Assessment*, 18(2), 192-203. doi: [10.1037/1040-3590.18.2.192](https://doi.org/10.1037/1040-3590.18.2.192)
- Drasgow, F., Levine, M. V., Tsien, S., Williams, B. A., & Mead, A. D. (1995). Fitting Polytomous Item Response Theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143-165. doi: [10.1177/014662169501900203](https://doi.org/10.1177/014662169501900203)
- Dumont, F. (2010). *A History of Personality Psychology: Theory, Science, and Research from Hellenism to the Twenty-first Century*. New York, NY: Cambridge University. doi: [10.1017/cbo9780511676093.002](https://doi.org/10.1017/cbo9780511676093.002)
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
- Eysenck, H. J. (1947). *Dimensions of Personality*. Londres, Reino Unido: Routledge & Kegan Paul.
- Eysenck, H. J. (1952). *The scientific study of personality*. Londres, Reino Unido: Routledge & Kegan Paul.
- Eysenck, H. J. (1958). A short questionnaire for the measurement of two dimensions of personality. *Journal of Applied Psychology*, 42(1), 14-17. doi: [10.1037/h0041738](https://doi.org/10.1037/h0041738)
- Eysenck, H. J., & Eysenck, S. B. G. (1969). *Personality structure and measurement*. Londres, Reino Unido: Routledge & Kegan Paul.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. Londres, Reino Unido: Hodder and Stoughton.
- Eysenck, H. J., Barret, P., Wilson, G. D., & Jackson, C. J. (1992). Primary trait measurement of the 21 components of the P-E-N System. *European Journal of Psychological Assessment*, 8(2), 109-117.
- Eysenck, H. J., Wilson, G. D., & Jackson, C. J. (1996). *Manual of the Eysenck Personality Profiler (Short)*. Guilford, Inglaterra: Psi-Press.
- Eysenck, S. B. G., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*, 54(4), 485-489. doi: [10.1016/j.paid.2012.09.022](https://doi.org/10.1016/j.paid.2012.09.022)
- Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism Scale. *Personality and Individual Differences*, 6(1), 21-29. doi: [10.1016/0191-8869\(85\)90026-1](https://doi.org/10.1016/0191-8869(85)90026-1)
- Flores-Mendoza, C., Ardila, R., Gallegos, M., Sampaio-Braga, L., Carvalhais-Santiago, B. M., & Andrade, D. M. (2016). Hans Eysenck in Latin America: His influence in the psychology, the study of personality and individual differences. *Personality and Individual Differences*, 103, 68-73. doi: [10.1016/j.paid.2016.04.049](https://doi.org/10.1016/j.paid.2016.04.049)
- Forbey, J. D., & Ben-Porath, Y. S. (2007). Computerized adaptive personality testing: A review and illustration

- with the MMPI-2 computerized adaptive version. *Psychological Assessment*, 19(1), 14-24. doi: [10.1037/1040-3590.19.1.14](https://doi.org/10.1037/1040-3590.19.1.14)
- Francis, L. J., Lewis, C. A., & Ziebertz, H. G. (2006). The Short-Form Revised Eysenck Personality Questionnaire (EPQR-S): A German edition. *Social Behavior Personality: An International Journal*, 34(2), 197-204. doi: [10.2224/sbp.2006.34.2.197](https://doi.org/10.2224/sbp.2006.34.2.197)
- Francis, L. J., Brown, L. B., & Philipchalk, R. (1992). The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the U.S.A. and Australia. *Personality and Individual Differences*, 13(4), 443-449. doi: [10.1016/0191-8869\(92\)90073-X](https://doi.org/10.1016/0191-8869(92)90073-X)
- Furnham, A., Eysenck, S. B. G., & Saklofske, D. H. (2008). The Eysenck personality measures: Fifty years of scale development. En G. J. Boyle, G. Matthews & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment* (Vol. II, pp. 199-218). Los Angeles, CA: Sage. doi: [10.4135/9781849200479.n10](https://doi.org/10.4135/9781849200479.n10)
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. En I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf (Eds.), *Personality Psychology in Europe*, (Vol. VII, pp. 7-28). Tilburg, The Netherlands: Tilburg University.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504-528. doi: [10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Guilford, J. P. (1940). *An inventory of factors STDCR*. California: Sheridan Supply.
- Hajek, A., Bock, J. O., & König, H. H. (2017). The role of personality in health care use: Results of a population-based longitudinal study in Germany. *PLoS One*, 12(7), 1-15. doi: [10.1371/journal.pone.0181716](https://doi.org/10.1371/journal.pone.0181716)
- Hengartner, M. P., Kawohl, W., Haker, H., Rössler, W., & Ajdacic-Gross, V. (2016). Big Five personality traits may inform public health policy and preventive medicine: Evidence from a cross-sectional and a prospective longitudinal epidemiologic study in a Swiss community. *Journal of Psychosomatic Research*, 84, 44-51. doi: [10.1016/j.jpsychores.2016.03.012](https://doi.org/10.1016/j.jpsychores.2016.03.012)
- Jackson, C. J., Furnham, A., Forde, L., & Cotter, T. (2000). The structure of the Eysenck Personality Profiler. *British Journal of Psychology*, 91(2), 223-239. doi: [10.1348/000712600161808](https://doi.org/10.1348/000712600161808)
- Jeronimus, B. F., Kotov, R., Riese, H., & Ormel, J. (2016). Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly decays with time: A meta-analysis on 59 longitudinal/prospective studies with 443 313 participants. *Psychological Medicine*, 46(14), 2883-2906. doi: [10.1017/S0033291716001653](https://doi.org/10.1017/S0033291716001653)
- Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 64(4), 241-256. doi: [10.1037/a0015309](https://doi.org/10.1037/a0015309)
- Lewis, C. A., Francis, L. J., Shevlin, N., & Forrest, S. (2002). Confirmatory factor analysis of the French translation of the Abbreviated Form of the Revised Eysenck Personality Questionnaire (EPQR-A). *European Journal of Psychological Assessment*, 18(2), 179-185. doi: [10.1027//1015-5759.18.2.179](https://doi.org/10.1027//1015-5759.18.2.179)
- Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T. K., Shirali, M. ... Deary, I. J. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*, 50(1), 6-11. doi: [10.1038/s41588-017-0013-8](https://doi.org/10.1038/s41588-017-0013-8)
- McCrae, R. R., & Costa, P. T. Jr. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- Milojev, P., Osborne, D., Greaves, L. M., Barlow, F. K., & Sibley, C. G. (2013). The Mini-IPIP6: Tiny yet highly stable markers of Big Six personality. *Journal of Research in Personality*, 47(6), 936-944. doi: [10.1016/j.jrp.2013.09.004](https://doi.org/10.1016/j.jrp.2013.09.004)
- Mitchell, R. L. C., & Kumari, V. (2016). Hans Eysenck's

- interface between the brain and personality: Modern evidence on the cognitive neuroscience of personality. *Personality and Individual Differences*, 103, 74-81. doi: [10.1016/j.paid.2016.04.009](https://doi.org/10.1016/j.paid.2016.04.009)
- Moeller, S. B., Bech, P., Kessing, L., Mortensen, E. L., Austin, S. F., & Bukh, J. D. (2015). A psychometric validation analysis of Eysenck's neuroticism and extraversion scales in a sample of first time depressed patients. *Journal of Depression and Anxiety*, 4(4), 1-6. doi: [10.4172/2167-1044.1000202](https://doi.org/10.4172/2167-1044.1000202)
- Muthén, L., & Muthén, B. (2010). *Mplus User's Guide*. (6aed.). Los Angeles, CA: Muthén & Muthén.
- Olea, J., & Ponsoda, V. (2013). *Tests adaptativos informatizados*. Madrid, España: UNED.
- Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., & Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck Personality Questionnaire - Short form (EPP-S). *Personality and Individual Differences*, 123, 76-80. doi: [10.1016/j.paid.2017.11.013](https://doi.org/10.1016/j.paid.2017.11.013)
- Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment*, 7(4), 347-364. doi: [10.1177/107319110000700404](https://doi.org/10.1177/107319110000700404)
- Reise, S. P., & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling Applications to Typical Performance Assessment*. Nueva York, NY: Routledge.
- Revelle, W. (2016). Hans Eysenck: Personality theorist. *Personality and Individual Differences*, 103, 32-39. doi: [10.1016/j.paid.2016.04.007](https://doi.org/10.1016/j.paid.2016.04.007)
- Rubio, V. J., Aguado, D., Hontangas, P. M., & Hernández, J. M. (2007). Psychometric properties of an emotional adjustment measure. An application of the Graded Response Model. *European Journal of Psychological Assessment*, 23(1), 39-46. doi: [10.1027/1015-5759.23.1.39](https://doi.org/10.1027/1015-5759.23.1.39)
- Salazar-Schneiderman, Tonetti, C. S., Sommers, L. M., & Greene, R. L. (2016). Advances in the use and interpretation of the MMPI-2. En U. Kumar (Ed.), *The Wiley Handbook of Personality Assessment* (pp. 158-172). Nueva York, NY: John Wiley & Sons. doi: [10.1002/9781119173489.ch12](https://doi.org/10.1002/9781119173489.ch12)
- Sánchez, R. O., & Ledesma, R. D. (2009). Análisis psicométrico del Inventario de Síntomas Revisado (SCL-90-r) en población clínica. *Revista Argentina de Clínica Psicológica*, 18(3), 265-274. Recuperado de <http://www.clinicapsicologica.org.ar>
- Sandín, B., Valiente, R. M., Chorot, P., Olmedo, M., & Santed, M. A. (2002). Versión española del cuestionario EPQR-Abreviado (EPQR-A) (I): Análisis exploratorio de la estructura factorial. *Revista de Psicopatología y Psicología Clínica*, 7(3), 195-205. doi: [10.5944/rppc.vol.7.num.3.2002.3933](https://doi.org/10.5944/rppc.vol.7.num.3.2002.3933)
- Sauer-Zavala, S., Wilner, J. G., & Barlow, D. H. (2017). Addressing neuroticism in psychological treatment. *Personality Disorders: Theory, Research, and Treatment*, 8(3), 191-198. doi: [10.1037/per0000224](https://doi.org/10.1037/per0000224)
- Sibley, C. G. (2012). The Mini-IPIP6: Item Response Theory analysis of a short measure of the Big-Six Factors of personality in New Zealand. *New Zealand Journal of Psychology*, 41(3), 20-31. Recuperado de <http://www.psychology.org.nz/publications-media/new-zealand-journal-of-psychology/#.W2mQkdVKjIV>
- Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E. S., Colodro-Conde, L., Ward, J., ... O'Donovan, M. C. (2016). Genome-wide analysis of over 106000 individuals identifies neuroticism-associated loci. *Molecular Psychiatry*, 21, 749-757. doi: [10.1038/mp.2016.49](https://doi.org/10.1038/mp.2016.49)
- Soto, C. J., & John, O. P. (2016). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117-143. doi: [10.1037/pspp0000096](https://doi.org/10.1037/pspp0000096)
- Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69-81. doi: [10.1016/j.jrp.2017.02.004](https://doi.org/10.1016/j.jrp.2017.02.004)
- Squillace, M., Picón-Janeiro, J., & Schmidt, V. (2013).

- Adaptación local del Cuestionario Revisado de Personalidad de Eysenck (Versión abreviada). *Revista Evaluar*, 13(1), 19-37. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Stark, S. (2001). MODFIT: A computer program for model-data fit. Manuscrito inédito, University of Illinois, Urbana-Champaign, IL, U.S.A.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251. doi: [10.1037/0033-2909.87.2.245](https://doi.org/10.1037/0033-2909.87.2.245)
- Tackett, J. L., & Lahey, B. B. (2017). Neuroticism. En T. A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model*. New York, NY: Oxford University. doi: [10.1093/oxfordhb/9780199352487.013.14](https://doi.org/10.1093/oxfordhb/9780199352487.013.14)
- Thissen, D. (2003). MULTILOG [software]. Chicago: Scientific Software International.
- Thurstone, L. L., & Thurstone, T. G. (1930). A neurotic inventory. *The Journal of Social Psychology*, 1(1), 3-30. doi: [10.1080/00224545.1930.9714128](https://doi.org/10.1080/00224545.1930.9714128)
- Tiwari, T., Singh, A. L., & Singh, I. L. (2009). The Short-Form Revised Eysenck Personality Questionnaire: A Hindi edition (EPQRS-H). *Industrial Psychiatry Journal*, 18(1), 27-31. doi: [10.4103/0972-6748.57854](https://doi.org/10.4103/0972-6748.57854)
- Vittengl, J. R. (2017). Who pays the price for high neuroticism? Moderators of longitudinal risks for depression and anxiety. *Psychological Medicine*, 47(10), 1794-1805. doi: [10.1017/S0033291717000253](https://doi.org/10.1017/S0033291717000253)
- Widiger, T. A. (2009). Neuroticism. En M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 129-146). New York, NY: Guilford.
- Widiger, T. A., & Oltmanns, J. R. (2017). Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry*, 16(2), 144-145. doi: [10.1002/wps.20411](https://doi.org/10.1002/wps.20411)
- Woodworth, R. S. (1919). Personal Data Sheet. Chicago, IL: C. H. Stoelting.
- Zambrano-Cruz, R. (2011). Revisión sistemática del Cuestionario de Personalidad de Eysenck (Eysenck Personality Questionnaire - EPQ). *Liberabit*, 17(2), 147-155. Recuperado de <http://revistaliberabit.com/es>
- Ziegler, M., Kemper, C. J., & Kruey, P. (2014). Short scales - Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, 35(4), 185-189. doi: [10.1027/1614-0001/a000148](https://doi.org/10.1027/1614-0001/a000148)
- Zuckerman, M. (2005). *Psychobiology of personality* (2^a ed.). New York, NY: Cambridge University. doi: [10.1017/cbo9780511813733](https://doi.org/10.1017/cbo9780511813733)

Motivational Regulation Strategies: A Questionnaire for its Measurement in Argentinian University Students

Estrategias de regulación motivacional: un cuestionario para su medición en estudiantes universitarios argentinos

Javier Sánchez-Rosas *¹, Rocío Soledad Aguirre²,
Nicolás Bovina-Martijena², Valeria Lucía Galarza²

1 - Laboratory of Psychological and Educational Assessment, Faculty of Psychology, National University of Cordoba, Argentina.

2 - Faculty of Psychology, National University of Cordoba, Argentina.

Recibido: 10/11/2018 Revisado: 28/12/2018 Aceptado: 02/02/2019

Introduction
Study 1
Method
Results
Study 2
Method
Results
Discussion
References

Abstract

The objective of this work was to evaluate the psychometric properties of a modified version of the Motivational Regulation Strategies Questionnaire of Wolters and Benzon (2013) in a sample of Argentinian university students. Thirty bilingual students answered the instrument in English and Spanish and the equivalence of the translation was analyzed. Spearman's correlations between the original and the translated version were strong (.71 to .90), which showed the equivalence. Then, the adapted questionnaire (N = 329) was administered and its internal structure and internal consistency were analyzed. The exploratory factor analysis showed a questionnaire consisting of 37 items and eight factors. The scales showed good internal consistency indexes ($\alpha = .68$ to $\alpha = .95$). The questionnaire evaluates value regulation, situational interest regulation, performance-approach regulation, performance-avoidance regulation, mastery-approach regulation, mastery-avoidance regulation, self-consequence, environmental structuring. The results and the need for new psychometric studies are discussed.

Keywords: *regulation of motivation, self-regulated learning, motivation, task value, interest, achievement goals, measurement, university students*

Resumen

El objetivo de este trabajo fue evaluar las propiedades psicométricas de una versión modificada del Cuestionario de Estrategias de Regulación Motivacional de Wolters y Benzon (2013) en una muestra de estudiantes universitarios argentinos. Treinta estudiantes bilingües respondieron al instrumento en inglés y español y se analizó la equivalencia de la traducción realizada. Las correlaciones de Spearman entre la versión original y la traducida fueron fuertes (.71 a .90), lo cual evidenció la equivalencia. Luego, se administró el cuestionario adaptado (N = 329) y se analizaron su estructura y consistencia internas. El análisis factorial exploratorio evidenció un cuestionario conformado por 37 ítems y ocho factores. Las escalas mostraron índices de consistencia interna buenos ($\alpha = .68$ a $\alpha = .95$). El cuestionario evalúa las regulaciones del valor, del interés situacional, de aproximación-rendimiento, de evitación-rendimiento, de aproximación-maestría, de evitación-maestría, auto-consecuencia y estructuración ambiental. Se discuten los resultados y la necesidad de nuevos estudios psicométricos.

Palabras clave: *regulación motivacional, aprendizaje auto-regulado, motivación, valor de la tarea, interés, metas de logro, medición, estudiantes universitarios*

Introduction

Most research on self-regulated learning has focused on cognitive and metacognitive strategies, and has barely addressed the way in which students regulate their motivation to cope with boredom, distraction or other motivational obstacles. Wolters (2003) has defined *motivational regulation* as the activity through which individuals decide to initiate, maintain or supplement their predisposition to start, focus on or complete a particular task or goal. It includes thoughts, actions or behaviors through which students act in order to influence their choices, their effort or persistence in academic tasks (Wolters, 2003). Some research (Grunschel, Schwinger, Steinmayr, & Fries, 2016; Ljubin-Golub, Petričević, & Rován, 2019; Schwinger & Otterpohl, 2017; Smit, de Brabander, Boekaerts, & Martens, 2017; Wolters & Benzon, 2013; Wolters & Rosenthal, 2000) has found that the students who most frequently use motivation regulation strategies have motivationally more adaptive beliefs (e.g., lower procrastination, higher task value and effort, higher levels of self-efficacy, and greater emphasis on mastery goals).

Researchers have designed instruments that evaluate a wide variety of motivational regulation strategies (Cabanach et al., 2009; Kim, Brady, & Wolters, 2018; Paulino, Sá, & Lopes da Silva, 2015; Schwinger, Steinmayr, & Spinath, 2009; Schwinger, von der Laden, & Spinath, 2007; Suárez-Riveiro & Fernández-Suárez, 2005; Suárez & Fernández 2011; Teng & Zhang, 2016; Wolters & Benzon, 2013) based on which they conducted psychometric studies that showed their suitability for measuring strategies of motivational regulation. In order to contribute to studies on motivational regulation, and in absence of an instrument validated in Argentina, this paper evaluates the psychometric properties of a modified

version of the Motivational Regulation Strategies Questionnaire of Wolters and Benzon (2013) for its use with university students.

Evaluation of motivational regulation strategies

Motivational regulation strategies are students' efforts to maintain or improve their motivation, which involve three phases necessary for effective and permanent self-regulation of motivation: knowledge, monitoring and control of motivation (Wolters & Benzon, 2013). The first phase involves meta-level understanding of students that reflects their knowledge or beliefs about motivation. The second phase requires monitoring of one's level or state of motivation. Finally, the third phase involves intentional actions in progress to intervene and control one's motivation, effort, or persistence.

Using an open survey, Wolters (1998, 2003) revealed more than a dozen different strategies for the regulation of motivation. Subsequent research has divided these strategies into five groups of particular importance: *environmental structuring*, *self-consequence*, *self-talk on performance* and *self-talk on mastery* (Cooper & Corpus, 2009).

In three subsequent studies, while replicating the findings of Wolters (Wolters, 1999; Wolters & Rosenthal, 2000), Schwinger et al. (2007) have also identified strategies for *environmental control*, *self-consequence*, *performance self-talk* and *master self-talk*. A scale of improvement on general interest consistent with Wolters (1999) was supported in one of these studies, but in the two remaining studies, it was divided into separate scales; one focused on maintaining situational interest and another focused on regulating personal value or importance (Wolters & Benzon, 2013). Based on the analysis of the open questions in the first of these studies, Schwinger et al. (2007) de-

vised some new items and found support for additional strategies based on performance-avoidance goals and the establishment of approach goals.

These studies have identified a wide variety of regulatory strategies, which led Schwinger et al. (2007) and Wolters and Benzon (2013) to construct separate instruments that measure the use of motivational regulation strategies. Next, we will focus on the description and proposal of adapting the Motivational Regulation Strategies Questionnaire of Wolters and Benzon (2013), which was facilitated by one of its authors, who in turn authorized its adaptation.

Motivational regulation strategies questionnaire

The Motivational Regulation Strategies Questionnaire (Wolters & Benzon, 2013) is a thirty-item instrument that measures six strategies of motivational regulation, and for which validity and reliability studies showed good psychometric properties. Specifically, evidence of internal structure was provided by exploratory factor analysis, criterion validity evidence through its relationships with motivational variables and internal consistency through Cronbach alpha coefficient.

These items were subject to exploratory analysis of principal components with oblimin rotation, providing strong support for the existence of six different types of strategies. These six factors together represented 69% of the total variance and the scales derived from each of them showed high reliability that went from $\alpha = .77$ to $\alpha = .91$. The bivariate correlations between the six strategies were positive and high ($r_s \geq .50$, $p_s < .01$), but not high enough so as to indicate that the strategies overlap in a single underlying construct. Finally, regarding criterion validity evidence, on the one hand, positive relationships were found between the six scales and task value, self-effi-

cacy, mastery goals, cognitive and metacognitive learning strategies and, on the other hand, negative relationships with procrastination and performance goals.

The set of strategies was constituted as follows: regulation of value, regulation of performance goals, self-consequence, environmental structuring, regulation of situational interest and regulation of mastery goals. The aim of *regulation of value* is to increase the value of the material to be studied or of the task at hand. To do this, the students focus on the relevance that the task to be completed may have for their personal lives and the usefulness of the materials. *Regulation of performance goals* encourages students to invoke their desire to do well on their exams or at classes in general, and to convince themselves to keep up their efforts and complete the task. *Self-consequence* is a prototypical way by which students regulate their motivation by self-managing the consequences of their own behavior, as well as identifying and administering extrinsic reinforcements or punishments to achieve particular objectives associated with the accomplishment of a task. *Environmental structuring* refers to students' efforts to adapt the study environment in order to carry out their tasks, as well as to adapt themselves physically. *Regulation of situational interest* is about students' attempts to complete their courses in a more pleasant, fun manner or as if it were a game. And finally, *regulation of mastery goals* is intended to fulfill the desire to improve their understanding or to learn as much as possible (Wolters & Benzon, 2013).

In the present work, a modified version of this instrument is adapted. Reference is made to a modified version in relation to the scales of regulation of achievement goals, given that the original instrument does not make a distinction of approach and avoidance between mastery goals and performance goals. Therefore, the version

was adapted using the scales of *regulation of value*, *self-consequence*, *environmental structuring* and *regulation of situational interest*, and was complemented with scales of regulation of *performance-approach goals*, *performance-avoidance goals*, *mastery-approach goals* and *mastery-avoidance goals*, which this investigation will expand on. Through two studies, we sought to

- (a) obtain a version translated into Spanish equivalent to the original version in English,
- (b) analyze the internal validity of the instrument through exploratory factor analysis, and
- (c) analyze the internal consistency through Cronbach's alpha coefficient.

Study 1: Scale adaptation

Method

Translation and equivalence analysis

In the present work, a direct translation from English into Spanish was carried out by an official translator and the authors of this work. When doing the translation, attention was directed to the semantic and syntactic properties, as well as to all the variations necessary so as to be applied to the target population without losing its meaning. Special attention was paid to the items having a clear, precise and simple formulation, trying to maintain the meaning of the construct the original item intended to measure.

To provide evidence on the equivalence of the translated version, an equivalence analysis was carried out between both versions of the instrument. Hambleton (1994) pointed out that one of the methods for carrying out this type of analysis is the implementation of the tests in the original language and translated into a sample of bilingual subjects. Subsequently, it is necessary to analyze if there is a correlation between the results obtained from the two versions.

Participants

The sample consisted of thirty bilingual students of the Faculty of Languages of the National University of Cordoba in the fourth year of undergraduate studying English translation, teaching or licentiate, 28 women, 20 to 29 years old.

Instrument

Scales adapted from the Motivational Regulation Strategies Questionnaire (Wolters & Bizon, 2013). For the measurement of motivational regulation strategies, four of the six scales of the Wolters and Bizon instrument (2013) were used in their original version. The students had to respond, in Spanish and English, to twenty items using a Likert scale ranging from 1 (*Never*) to 5 (*Always*). The scales employed were: regulation of value (e.g., *I make an effort to relate what we're learning to my personal interests*; $\alpha = .91$, 6 items), self-consequence (e.g., *I promise myself I can do something I want later if I finish the assigned work now*; $\alpha = .91$, 5 items), environmental structuring (e.g., *I change my surroundings so that it is easy to concentrate on the work*; $\alpha = .77$, 4 items) and regulation of the situational interest (e.g., *I make studying more enjoyable by turning it into a game*; $\alpha = .88$, 5 items).

Procedure and data analysis

The English version of the questionnaire was administered at the Faculty of Languages to thirty bilingual students. They were informed that participation was voluntary and that confidential treatment of the information was guaranteed. After one week, the Spanish version was administered to the same thirty students. The data was

Table 1

Descriptive statistics of the scales and their correlations.

Scale	Spearman coefficient	M Original	DS Original	M Adaptationn	DS Adaptation
regulation of value	.82**	3.40	0.75	3.24	0.81
self-consequence	.90**	3.07	1.06	3.00	1.15
regulation of situational interest	.85**	2.52	0.92	2.27	0.80
environmental structuring	.71**	3.04	0.64	2.78	0.74

Note. ** $p < .01$.

loaded and analyzed with SPSS 24 (IBM, 2016).

First, the total score of the scales was obtained (value regulation, self-consequence, situational interest regulation, environmental structuring) by adding and calculating the average of the individual items. Then, the analysis of the correlations between each scale of the original instrument with its adapted version was performed, using the Spearman correlation coefficient. According to Martínez-Ortega, Tuya-Pendás, Martínez-Ortega, Pérez-Abreu and Cánovas (2009), Spearman's correlations can be interpreted as scarce or none (0 to .25), weak (.26 to .50), moderate and high (.51 to .75), and between high and perfect (.76 to 1).

Results

The correlations between the original scales and the adapted scales were moderate and high. Table 1 shows the descriptive statistics of the scales and their correlations.

Study 2: Psychometric properties of the modified version of the Motivational Regulation Strategies Questionnaire Method

Participants

The sample consisted in 329 university students of the Faculty of Psychology (49%) and of the Faculty of Exact, Physical and Natural Sciences (51%) of the National University of Córdoba, selected by accidental non-probabilistic sampling. The participants were between 18 and 59 years old; 63% of the sample consisted of women.

Instruments

An important modification made to the original version was the discrimination of the dimensions of approach and avoidance in the strategies of regulation of mastery goals and performance goals and, therefore, the creation of new items to measure these dimensions. In writing the items of the new strategies, the format used and recommended by Elliot and Murayama (2008) for the measurement of achievement goals was respected.

Motivational Regulation Strategies Questionnaire. The total set of strategies to be evaluated was made up of eight scales in the following fashion: regulation of value (e.g., *I think up situations where it would be helpful for me to know the material or skills*), self-consequence (e.g., *I promise myself some type of a reward if I get my readings or studying done*), environmental structuring (e.g., *I try to get rid of any distractions that are around me*), regulation of situational interest (e.g., *I make doing the work seem enjoyable by focusing on something about it that is fun*), regulation of performance-approach goals (e.g., *I remind myself that my goal is to do better than others at exams and tasks*), regulation of performance-avoidance goals (e.g., *I remind myself that my goal is to avoid performing worse than other students*), regulation of mastery-approach goals (e.g., *I tell myself I should keep working just to learn as much as I can*), regulation of mastery-avoidance goals (e.g., *I am persuaded to continue studying because my goal is to avoid learning less than I could*).

In turn, the response scale of 7 (original version) was modified to 5 points by replacing the expression *not at all* with *never*, and *very often* with *always*. Consequently, the students had to respond using a Likert scale that ranges from 1 (*Never*) to 5 (*Always*) to indicate how often they used each strategy.

Procedure

The instrument was administered through the online survey system LimeSurvey (Pérez, 2007) due to its practicality for this type of surveys. The data obtained were loaded into the statistical program IBM SPSS versión 24.0 (IBM Corporation, 2016), where the necessary statistical procedures were carried out to complete the objectives set out in this work.

Data Analysis

Prior to the central analysis, an initial exploration of all the items was carried out in order to evaluate the missing values, the univariate and multivariate atypical cases and the normal distribution (George & Mallery, 2007). Atypical cases were identified by calculating the z score for each variable (z score > 3.29 was considered atypical) and Mahalanobis distance ($p < .001$ was considered atypical).

An exploratory factorial analysis was carried out in order to identify the constructs that underlie the items of motivational regulation strategies. The factorial analysis guidelines recommended by Fabrigar, Wegener, MacCallum, and Strahan (1999) were followed. The maximum likelihood factor extraction method was used (Jöreskog & Lawley, 1968) because it is believed to produce the best estimates of the parameters (Pedhazur, 1982). Since the strategies are interrelated, an oblique rotation method (promax) was used. Multiple criteria were used for the selection of the number of factors: (a) the Kaiser rule of eigenvalues greater than 1 (Kaiser, 1960), (b) the sedimentation graph (Cattell, 1966), and (c) the interpretation of the rotated factors.

Then the internal consistency was estimated by means of Cronbach's alpha coefficient, revealing to what extent the different items of a test measure the same variable. The criterion used was the following: .70 acceptable, .80 good, and .90 excellent (George & Mallery, 2007).

Results

Once the adequacy of the collected data was revised, the descriptive statistics of mean, standard deviation and the skewness and kurtosis indices of each variable were calculated (Table 2).

Table 2
Descriptive statistics.

Scale	Items	Minimum	Maximum	M	SD	Skewness	Kurtosis
Regulation of value	1	1	5	3.65	0.93	-0.45	-0.04
	2	1	5	3.71	0.92	-0.46	-0.04
	3	1	5	3.77	0.96	-0.54	-0.16
	4	1	5	3.57	0.96	-0.37	-0.25
	5	1	5	3.64	1.07	-0.53	-0.45
	6	1	5	3.53	0.99	-0.44	-0.20
Self-Consequence	1	1	5	2.69	1.33	0.25	-1.13
	2	1	5	3.00	1.27	-0.05	-1.02
	3	1	5	3.14	1.22	-0.24	-0.95
	4	1	5	2.74	1.24	0.20	-0.96
	5	1	5	2.99	1.22	-0.11	-0.96
Environmental Structuring	1	1	5	3.36	1.03	-0.17	-0.66
	2	1	5	3.35	1.09	-0.14	-0.81
	3	1	5	3.19	1.18	-0.23	-0.84
	4	1	5	3.70	1.15	-0.77	-0.14
Regulation of situational Interest	1	1	5	1.76	0.97	1.23	0.93
	2	1	5	1.79	0.99	1.19	0.73
	3	1	5	2.15	0.99	0.51	-0.45
	4	1	5	2.31	1.01	0.38	-0.48
	5	1	5	2.72	1.01	-0.03	-0.46
Regulation of Performance-approach Goals	1	1	5	2.08	1.26	0.88	-0.41
	2	1	5	1.98	1.13	1.03	0.15
	3	1	5	1.79	1.14	1.37	0.82
	4	1	5	1.98	1.14	1.02	0.11
	5	1	5	1.71	1.09	1.51	1.37
Regulation of Performance-avoidance Goals	1	1	5	2.04	1.23	0.95	-0.18
	2	1	5	1.93	1.11	1.08	0.34
	3	1	5	1.83	1.17	1.34	0.79
	4	1	5	1.88	1.14	1.19	0.43
	5	1	5	1.81	1.10	1.34	1.01
Regulation of Mastery-approach Goals	1	1	5	3.89	1.03	-0.79	0.18
	2	1	5	3.91	1.00	-0.71	0.01
	3	1	5	4.04	0.92	-0.83	0.48
	4	1	5	3.90	0.96	-0.45	-0.54
	5	1	5	3.91	0.98	-0.73	0.16
Regulation of Mastery-avoidance Goals	1	1	5	3.11	1.29	-0.21	-0.97
	2	1	5	3.06	1.28	-0.19	-0.94
	3	1	5	3.29	1.29	-0.50	-0.76
	4	1	5	3.06	1.30	-0.20	-0.99
	5	1	5	3.12	1.33	-0.31	-1.01

In general, the items presented values within the normal parameters since no skewness and kurtosis value greater than ± 2 were observed.

Analysis of internal structure

The exploratory factor analysis permitted the identification of eight factors that underlie the

items of motivational regulation strategies. Table 3 shows the results of the exploratory factor analysis, the eigenvalues, variance percentages and factor loads.

Table 4 shows the descriptive statistics for each scale and the correlations between them.

Table 3
Exploratory factor analysis of the items of motivational regulation.

Ítems	1	2	3	4	5	6	7	8
2. I think about reading or studying what I was assigned because I am trying to avoid having a worse performance than other students.	.95							
4. I think about making an effort because my goal is to avoid having a worse performance than other students.	.92							
5. I tell myself that I must continue studying because my goal is to prevent myself from doing worse than others in this area.	.87							
1. I remind myself that my goal is to avoid doing worse than other students.	.85							
3. I remind myself that my goal is to avoid doing worse than others at exams and tasks of this subject.	.83							
2. I think I should keep working because my goal is to avoid learning less than I could.		.92						
4. I am convinced of working hard because my goal is to avoid learning less than is possible.		.90						
3. I challenge myself to complete the work because my goal is to avoid not understanding of the material.		.86						
5. I tell myself that I should study since my goal is to avoid not mastering a large part of the material.		.79						
1. I am persuaded to continue studying because my goal is to avoid learning less than I could.		.75						
2. I make a deal with myself that if I get a certain amount of the work done I can do something fun afterwards.			.90					
3. I tell myself I can do something I like later if right now I do the work I have to get done.			.86					
5. I promise myself I can do something I want later if I finish the assigned work now.			.84					
4. I set a goal for how much I need to study and promise myself a reward if I reach that goal.			.82					
1. I promise myself some kind of a reward if I get my readings or studying done.			.76					

Ítems	1	2	3	4	5	6	7	8
3. I challenge myself to complete the work because my goal is to understand most of the material.				.86				
4. I am convinced of working hard because my goal is to learn as much as possible.				.82				
2. I think I should keep working because my goal is to learn as much as I can.				.76				
1. I am persuaded to continue studying because my goal is to learn as much as I can.				.74				
5. I tell myself that I should study since my goal is to master most of the material.				.69				
2. I think about reading or studying what I was assigned because I am trying to perform better than other students.					.92			
1. I remind myself that my goal is to perform better than other students.					.90			
3. I remind myself that my goal is to do better than others in exams and homework assignments.					.80			
4. I think about making an effort since my goal is to have a good performance in relation to other students.				.61				
5. I tell myself that I must continue studying because my goal is to do better than others in this area.					.59			
1. I make studying more enjoyable by turning it into a game.						.94		
2. I try to make a game out of learning the material or completing the assignment.						.91		
3. I make doing the work enjoyable by focusing on something about it that is fun.						.68		
4. I try to get myself to see how doing the work can be fun.						.61		
5. I think of a way to make the work seem enjoyable to complete.						.45		
2. I make sure I have as few distractions as possible.							.93	
1. I try to get rid of any distractions that are around me.							.78	
3. I change my surroundings so that it is easy to concentrate on the work.							.53	
4. I try to connect the material with something I like doing or find interesting,								.73
3. I make an effort to relate what we're learning to my personal interests.								.67
2. I try to make the material seem more useful by relating it to what I want to do in my life.								.59
1. I think up situations where it would be helpful for me to know the material or the skills.								.32
Eigenvalue	7.85	5.38	3.94	2.65	2.40	2.03	1.46	1.15
Percentage of variance explained (72.63%)	21.23	14.55	10.64	7.16	6.50	5.48	3.95	3.12

Note. Item loadings below .30 are not shown. 1 = regulation of performance-avoidance goals, 2 = regulation of mastery-avoidance goals, 3 = self-consequence, 4 = regulation of mastery-approach goals, 5 = regulation of performance-approach goals, 6 = regulation of situational interest, 7 = environmental structuring, 8 = regulation of value. * Items 5 and 6 of regulation of value and item 4 of environmental structuring were eliminated due to scores < .30.

Table 4
Descriptive statistics and bivariate correlations of the scales.

	1	2	3	4	5	6	7	8
1 Reg. of Value	-							
2 Self-Consequence	.08	-						
3 Environmental Structuring	.09	.14*	-					
4 Reg. of Situational Interest	.16**	.26**	.12*	-				
5 Reg. of Performance-approach Goals	.04	.18**	.12*	.02	-			
6 Reg. of Performance-avoidance Goals	.02	.21**	.06	.09	.65**	-		
7 Reg. of Mastery-approach Goals	.36**	.09	.16**	.05	.06	.02	-	
8 Reg. of Mastery-avoidance Goals	.18**	.18**	.11*	.12*	.08	.26**	.49**	-
M	3.68	2.91	3.40	2.15	1.91	1.90	3.93	3.13
SD	0.67	1.10	0.82	0.80	1.00	1.05	0.81	1.15
Skewness	-0.37	-0.01	-0.13	0.70	1.15	1.17	-0.51	-0.30
Kurtosis	-0.07	-0.93	-0.48	0.29	0.63	0.61	-0.10	-0.67

Note. * $p < .05$ ** $p < .01$.

Reliability analysis

Then, we can see the Cronbach's α score obtained for the different scales that make up the motivational regulation strategy questionnaire and how that coefficient varies if any of its items is excluded (Table 5).

As Table 5 shows, all the scales presented values between *acceptable*, *good* and *excellent* internal consistency (George & Mallery, 2007), except for regulation of value whose marginally acceptable value ($\alpha = .68$) can be considered adequate for research purposes. In addition, positive and high correlations are observed between the items and the total.

Discussion

The objective of this work was to evaluate the psychometric properties of a modified version of the questionnaire of motivational regulation

strategies for use with Argentine university students. A direct translation of the original instrument from English to Spanish was carried out, its internal structure was analyzed through exploratory factor analysis, and its internal consistency was evaluated using the Cronbach's alpha coefficient. The objectives proposed were satisfactorily fulfilled, obtaining a questionnaire of 37 items, grouped into eight scales with adequate indices of internal consistency. Next, the results obtained are discussed.

Equivalence analysis

First, four of the six scales (regulation of value, self-consequence, environmental structuring and regulation of situational interest) developed by Wolters and Benzon (2013) were translated. In order to evaluate the equivalence between the English and Spanish versions of the selected scales, these versions were applied to a bilingual sam-

Table 5
Item-Total Statistics.

Scale	Items	Corrected Item-Total Correlation	Cronbach's α if Item Deleted	Cronbach's α
Regulation of value	1	.34	.69	.68
	2	.51	.58	
	3	.48	.60	
	4	.51	.57	
Self-Consequence	1	.75	.91	.92
	2	.84	.89	
	3	.79	.90	
	4	.79	.90	
	5	.80	.90	
Environmental Structuring	1	.59	.65	.78
	2	.67	.60	
	3	.52	.84	
Regulation of Situational Interest	1	.74	.81	.86
	2	.72	.82	
	3	.68	.83	
	4	.69	.83	
	5	.55	.86	
Regulation of Performance-Approach Goals	1	.81	.90	.92
	2	.80	.90	
	3	.83	.89	
	4	.78	.91	
	5	.75	.91	
Regulation of Performance-Avoidance Goals	1	.83	.94	.95
	2	.88	.93	
	3	.85	.94	
	4	.89	.93	
	5	.86	.94	
Regulation of Mastery-Approach Goals	1	.71	.87	.89
	2	.76	.85	
	3	.76	.86	
	4	.76	.85	
	5	.65	.88	
Regulation of Mastery-Avoidance Goals	1	.75	.92	.93
	2	.85	.90	
	3	.82	.91	
	4	.85	.90	
	5	.78	.92	

ple. Subsequently, Spearman correlations were estimated with satisfactory results. The high correlations obtained between the variables of both versions (ρ between .71 and .90) confirm that the original and the translated version are equivalent.

Internal structure

The two non-selected scales of the [Wolters and Benzon \(2013\)](#) instrument assess the regulation of goals based on the dichotomous goal model ([Maehr & Nicholls, 1980](#)). Currently, the 2 x 2 model of achievement goals ([Elliot, 1999](#)) has a more complex development of goals that includes the approach-avoidance dimensions in the goals of mastery and performance. Accordingly, it was decided to develop items that would allow for the estimation of the regulation of the four goals of the 2 x 2 model of achievement goals ([Elliot, 1999](#)); and consequently, the internal structure of a set of items that sought to measure eight strategies of motivational regulation (regulation of value, self-consequence, environmental structuring, regulation of situational interest, regulation of mastery-approach goals, regulation of mastery-avoidance goals, regulation of performance-approach goals, regulation of performance-avoidance goals).

The internal structure of the instrument was made up of the eight expected factors and 37 items. Items 5 and 6 of *regulation of value* and 4 of *environmental structuring* presented low correlations with their respective factor, so they were eliminated, improving the factorial structure of the instrument. In the original instrument, these items also charged lower than the rest of the items in their respective scales. The value regulation strategies point to the link of the task with a personal aspect (personal goals, usefulness for the fulfillment of personal achievements). The con-

tent of item 5 of regulation of value does not explicitly show this link (*I tell myself that it is important to learn the material because I will need it later in life*) and this could be the reason for the low factorial load (.28). On the other hand, item 6 of *regulation of value* (*I try to make myself see how knowing the material is personally relevant*) presented a double factorial load on the factors regulation of value and regulation of mastery goals. The complexity of the writing of the item would lie in the fact that the act of knowing is a central aspect of the mastery goals while being relevant to the importance value of the learner. On the other hand, the low factorial load of item 4 of environmental structuring (*I try to study at a time when I can be more focused*) could be due to the fact that the regulation of the schedule does not necessarily imply regulation of the physical environment (visual, auditory stimuli). Added to this, the regulation of the environment may be more likely than the schedule, which is more subject to hourly availability, so that its covariation may be low.

Internal consistency

The results of internal consistency, evaluated by Cronbach's alpha coefficient, were good ([George & Mallery, 2007](#)) and similar to those obtained in the original instrument (between .78 and .95). The scale of *regulation of value* showed a marginally acceptable value ($\alpha = .68$), probably attributable to the small number of items. Further studies should consider the writing of additional and homogeneous items for an estimate that improves the internal consistency of the scale. In spite of this, the value obtained from internal consistency for the scale of regulation of value can be considered adequate for research purposes.

Limitations and future investigation

There is a wide variety of motivational and affective constructs related to learning, performance (Eccles, 2005; Pekrun, 2006; Pintrich, 2003) and to which motivational regulation strategies would be related. Among these constructs are task value (Eccles, 2005), achievement goals (Elliott & Murayama, 2008) and enjoyment (Pekrun, 2006). Considering that there are validated instruments for the evaluation of these constructs in Argentine university students, evidence could be provided of the relationship between the measurements made by the scales of this study and those made by the instruments that assess task value (Sánchez-Rosas, Lou, Lin, & Larroza, 2017), achievement goals (Sánchez-Rosas, 2015a) and enjoyment (Sánchez-Rosas, 2015b).

On the other hand, it is important to mention that the sample was limited to university students of the Faculty of Psychology and the Faculty of Physical and Natural Sciences, 63% women and 37% men. In future studies, it is suggested that the use of motivation regulation strategies be evaluated in students of other careers, and the possible differences be analyzed, in addition to conducting a study of gender differences in the use of the strategies. In turn, the same study could be conducted focusing on students of lower grades, since, in contrast, university students may be able to differentiate between concepts such as importance and interest and, therefore, distinguish the use of strategies designed to regulate every aspect of motivation.

The instrument studied here cannot be considered as an evaluation of all possible types of motivational regulation strategies. It is evident that there are important aspects of student motivation that are not well represented by the eight types of strategies evaluated with this instrument. For example, none of the strategies clearly reflects

the efforts of students to strengthen or maintain their perceived competence or self-efficacy to complete their academic work. However, self-efficacy is a critical aspect of motivation and previous work has found some evidence of this type of regulatory strategy (McCann & Turner, 2004, Wolters, 1998). Additional studies designed to discover and evaluate new strategies associated with these other forms of motivation represent another useful line of research.

On the other hand, in relation to the administered protocol, it is believed that the participants at the time of responding may not recognize differences between the items, due to the similarity in their content, as they are arranged according to the strategies. This could be considered a bias when measuring the strategies used. In future research, it is suggested that the items of different scales be interspersed and evaluated if there is a change.

The internal structure of the questionnaire is subject to the completion of a confirmatory factor analysis that provides evidence in favor of a model with the eight motivational regulation strategies identified in this exploratory study. As such, given the factorial loads and internal consistency values, it would be interesting to test the structure of an instrument with fewer items.

The availability of this instrument will lead to the measurement of motivational regulation in our context. For example, researchers interested in investigating the effects of programs aimed at developing self-regulated learning strategies, specifically motivational regulation, could perform pre-post measurements with this instrument. Or, how these strategies relate to other variables relevant to learning and academic performance could be explored.

References

- Cabanach, R. G., Valle, A., Gerpe, M. G., Rodríguez, S., Piñeiro, I., & Rosário, P. (2009). Diseño y validación de un cuestionario de gestión motivacional. *Revista de Psicodidáctica*, 14(1), 29-47. Recuperado de <http://www.chu.es/ojs/index.php/psicodidactica>
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245-276. doi: 10.1207/s15327906mbr0102_10
- Cooper, C. A., & Corpus, J. H. (2009). Learners' developing knowledge of strategies for regulating motivation. *Journal of Applied Developmental Psychology*, 30(4), 525-536. doi: 10.1016/j.appdev.2008.12.032
- Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York: Guilford Press.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169-189. doi: 10.1207/s15326985ep3403_3
- Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, 100(3), 613-628. doi: 10.1037/0022-0663.100.3.613
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. doi: 10.1037//1082-989x.4.3.272
- George, D., & Mallery, P. (2007). *SPSS for Windows: Step by step 14.0 update* (7th ed.). Boston, MA: Allyn & Bacon.
- Grunschel, C., Schwinger, M., Steinmayr, R., & Fries, S. (2016). Effects of using motivational regulation strategies on students' academic procrastination, academic performance, and well-being. *Learning and Individual Differences*, 49, 162-170. doi: 10.1016/j.lindif.2016.06.008
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-244.
- IBM Corporation (2016). IBM SPSS Statistics for Windows, Version 24.0 [Software de cómputo]. Armonk, NY: IBM.
- Jöreskog, K. G., & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 21(1), 85-96. doi: 10.1111/j.2044-8317.1968.tb00399.x
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151. doi: 10.1177/001316446002000116
- Kim, Y., Brady, A. C., & Wolters, C. A. (2018). Development and validation of the brief regulation of motivation scale. *Learning and Individual Differences*, 67, 259-265. doi: 10.1016/j.lindif.2017.12.010
- Ljubin-Golub, T., Petričević, E., & Rován, D. (2019). The role of personality in motivational regulation and academic procrastination. *Educational Psychology*, 1-19. doi: 10.1080/01443410.2018.1537479
- Maehr, M. L., & Nicholls, J. G. (1980). Culture and achievement motivation: A second look. In N. Warren (Ed.), *Studies in Cross-Cultural Psychology*, (pp. 221-267), New York: Academic Press.
- Martínez-Ortega, R. M., Tuya-Pendás, L. C., Martínez-Ortega, M., Pérez-Abreu, A., & Cánovas, A. M. (2009). El coeficiente de correlación de los rangos de Spearman, caracterización. *Revista Habanera de Ciencias Médicas*, 8(2), 1-19. Recuperado de <http://www.revhabanera.sld.cu/index.php/rhab>
- McCann, E. J., & Turner, J. E. (2004). Increasing student learning through volitional control. *The Teachers College Record*, 106(9), 1695-1714. doi: 10.1111/j.1467-9620.2004.00401.x
- Paulino, P., Sá, I., & Lopes da Silva, A. (2015). Crenças e estratégias da motivação na aprendizagem: Desenvolvimento de uma escala. *Psychologica*, 58(1), 65-

87. doi: [10.14195/1647-8606_58-1_4](https://doi.org/10.14195/1647-8606_58-1_4)
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd. ed.). Orlando, FL: Harcourt Brace.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315-341. doi: [10.1007/s10648-006-9029-9](https://doi.org/10.1007/s10648-006-9029-9)
- Pérez, C. J. M. (2007). *Manual de Usuario de la plataforma de encuestas en línea: LimeSurvey*. Versión 1.0, Licencia de Documentación Libre GNU.
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667-686. doi: [10.1037/0022-0663.95.4.667](https://doi.org/10.1037/0022-0663.95.4.667)
- Sánchez-Rosas, J. (2015a). Validation of the Achievement Goal Questionnaire-Revised in Argentinean university students (A-AGQ-R). *International Journal of Psychological Research*, 8(1), 10-23. doi: [10.21500/20112084.641](https://doi.org/10.21500/20112084.641)
- Sánchez-Rosas, J. (2015b). The Achievement Emotions Questionnaire-Argentine (AEQ-AR): Internal and external validity, reliability, gender differences and norm-referenced interpretation of test scores. *Revista Evaluar*, 15(1), 41-74. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Sánchez-Rosas, J., Lou, Y. C., Lin, H. F., & Larroza, S. (2017). A Spanish version of the Achievement Task Value Scale for University Students: Internal, convergent, and criterion validity and reliability in Argentinian students. *Pensando Psicología*, 13(21), 41-57. doi: [10.16925/pe.v13i21.1713](https://doi.org/10.16925/pe.v13i21.1713)
- Schwinger, M., & Otterpohl, N. (2017). Which one works best? Considering the relative importance of motivational regulation strategies. *Learning and Individual Differences*, 53, 122-132. doi: [10.1016/j.lindif.2016.12.003](https://doi.org/10.1016/j.lindif.2016.12.003)
- Schwinger, M., Steinmayr, R., & Spinath, B. (2009). How do motivational regulation strategies affect achievement: Mediated by effort management and moderated by intelligence. *Learning and Individual Differences*, 19(4), 621-627. doi: [10.1016/j.lindif.2009.08.006](https://doi.org/10.1016/j.lindif.2009.08.006)
- Schwinger, M., Von der Laden, T., & Spinath, B. (2007). Strategien zur Motivationsregulation und ihre Erfassung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39(2), 57-69. doi: [10.1026/0049-8637.39.2.57](https://doi.org/10.1026/0049-8637.39.2.57)
- Smit, K., de Brabander, C. J., Boekaerts, M., & Martens, R. L. (2017). The self-regulation of motivation: Motivational strategies as mediator between motivational beliefs and engagement for learning. *International Journal of Educational Research*, 82, 124-134. doi: [10.1016/j.ijer.2017.01.006](https://doi.org/10.1016/j.ijer.2017.01.006)
- Suárez-Riveiro, J. M., & Fernández-Suárez, A. P. (2005). Escalas de evaluación de las estrategias motivacionales de los estudiantes. *Anales de Psicología*, 21(1), 116-128. Recuperado de <http://revistas.um.es/analesps>
- Suárez, J. M., & Fernández, A. P. (2011). Evaluación de las estrategias de autorregulación afectivo-motivacional de los estudiantes: Las EEMA-VS. *Anales de Psicología*, 27(2), 369-380. Recuperado de <http://revistas.um.es/analesps>
- Teng, L. S., & Zhang, L. J. (2016). Fostering strategic learning: The development and validation of the Writing Strategies for Motivational Regulation Questionnaire (WSMRQ). *The Asia-Pacific Education Researcher*, 25(1), 123-134. doi: [10.1007/s40299-015-0243-4](https://doi.org/10.1007/s40299-015-0243-4)
- Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology*, 90(2), 224-235. doi: [10.1037/0022-0663.90.2.224](https://doi.org/10.1037/0022-0663.90.2.224)
- Wolters, C. A. (1999). The relation between high school students' motivational regulation and their use of learning strategies, effort, and classroom performance. *Learning and Individual Differences*, 11(3), 281-299. doi: [10.1016/s1041-6080\(99\)80004-1](https://doi.org/10.1016/s1041-6080(99)80004-1)
- Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learn-

ing. *Educational Psychologist*, 38(4), 189-205. doi:
[10.1207/s15326985ep3804_1](https://doi.org/10.1207/s15326985ep3804_1)

Wolters, C. A., & Benzon, M. B. (2013). Assessing and predicting college students' use of strategies for the self-regulation of motivation. *The Journal of Experimental Education*, 81(2), 199-221. doi:
[10.1080/00220973.2012.699901](https://doi.org/10.1080/00220973.2012.699901)

Wolters, C. A., & Rosenthal, H. (2000). The relation between students' motivational beliefs and their use of motivational regulation strategies. *International Journal of Educational Research*, 33(7-8), 801-820. doi: [10.1016/s0883-0355\(00\)00051-3](https://doi.org/10.1016/s0883-0355(00)00051-3)

La evaluación multidimensional de la ansiedad: Técnicas, aplicaciones y alcances

The multidimensional evaluation of anxiety: Techniques, applications and scope

Iliana Díaz-Kuaik * ¹, Guadalupe de la Iglesia ²

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

2 - Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad de Palermo, Universidad de Buenos Aires, Argentina.

Introducción

Evaluación de las dimensiones

Aplicaciones clínicas de la evaluación multidimensional de la ansiedad

Alcances de la evaluación multidimensional en investigación clínica

Conclusiones

Referencias

Recibido: 06/09/2018 Revisado: 08/10/2018 Aceptado: 05/11/2018

Resumen

La ansiedad es un constructo complejo que requiere la valoración de múltiples dimensiones. Conforme fue creciendo el conocimiento sobre los mecanismos implicados en la respuesta de ansiedad, se fue diversificando la tecnología utilizada para su evaluación. El objetivo de este trabajo es realizar un recorrido instrumental por la evaluación de las distintas dimensiones, definidas según la teoría cognitivo-conductual de la ansiedad (Clark & Beck, 2012). Con tal propósito, se exponen las principales técnicas utilizadas y las propuestas instrumentales generadas por los diferentes modelos teóricos que abordan la ansiedad con sus ventajas y limitaciones. Asimismo, se analizan las aplicaciones clínicas de la evaluación multidimensional de la ansiedad, así como sus alcances en investigación clínica.

Palabras clave: *ansiedad, evaluación multidimensional, aplicaciones clínicas, limitaciones*

Summary

Anxiety is a complex construct that requires the assessment of multiple dimensions. As the body of knowledge about the mechanisms involved in the anxiety response increased, the technology used for its evaluation was diversified. The objective of this work is to carry out an instrumental journey through the evaluation of the different dimensions, defined in light of the cognitive behavioral theory of anxiety (Clark & Beck, 2012). With this purpose, the main techniques used and the instrumental proposals generated by the different theoretical models that address anxiety are exposed along with their advantages and limitations. In addition, the clinical applications of the multidimensional evaluation of anxiety, as well as its scope in clinical research, are analyzed.

Keywords: *anxiety, multidimensional evaluation, clinical applications, limitations*

Introducción

La ansiedad es un constructo complejo abordado por múltiples modelos teóricos que ha evolucionado junto al desarrollo de la disciplina psicológica (De Ansorena-Cao, Cobo-Reinoso, & Romero-Cagigal, 1983; González-Martínez, 1993; Sierra, Ortega, & Zubeidat, 2003). Dado que la conceptualización actual de la ansiedad es multidimensional, su operacionalización demanda la valoración de cuatro sistemas de respuesta: fisiológico, cognitivo, conductual y afectivo (Clark & Beck, 2012; González-Martínez, 1993). La dimensión *fisiológica* alude a las respuestas defensivas automáticas que preparan al organismo para lidiar con circunstancias aversivas. La dimensión *cognitiva* refiere al procesamiento cognitivo a través del cual se releva y analiza información relevante sobre la situación, los recursos y el resultado esperado. La dimensión *conductual* refleja las respuestas conductuales llevadas a cabo para prevenir, controlar, luchar o huir del peligro. La dimensión *afectiva* alude al malestar subjetivo generado por la experiencia ansiosa (Clark & Beck, 2012). Para evaluar estas dimensiones se dispone de una amplia variedad de técnicas y propuestas instrumentales generadas por los diversos modelos teóricos. Este trabajo tiene como objetivo describir dichas técnicas y propuestas, sus principales ventajas y limitaciones, considerando las aplicaciones clínicas que posee la evaluación multidimensional y sus alcances en investigación.

Evaluación de la dimensión fisiológica

La evaluación de las variables *fisiológicas* de la ansiedad es un legado de los primeros modelos científicos de las emociones aportados por las teorías psicofisiológicas (James 1884, 1890; Lange, 1885). Las técnicas utilizadas en las últimas déca-

das difieren sustancialmente de las propuestas instrumentales originales. Sobre todo, porque actualmente existen medidores fisiológicos portables o se prefieren métodos indirectos de evaluación, tales como la observación y los autoinformes, con ciertas ventajas sobre las técnicas objetivas o directas (Buela-Casal & Sierra, 2004; Calero, Márquez, Vizcarro, & Fernández-Ballesteros, 2013; Carrobes & Buela-Casal, 1997).

Evaluación directa de las variables fisiológicas

La evaluación directa posee una larga tradición de la mano de la psicofisiología y de la psicología experimental, de gran incidencia para el campo de la psicología clínica y de la salud. Ha consistido en la valoración de ciertas respuestas fisiológicas consideradas como índices de procesos o estados psicológicos (Vila, 2004). Los autores de los primeros modelos de las emociones dedicaron especial atención a las respuestas fisiológicas vinculadas a los procesos afectivos y emocionales (James, 1884, 1890; Lange, 1885). Se interesaron en la activación del sistema nervioso autónomo y somático y propusieron la existencia de cierta especificidad fisiológica de las emociones. Posteriormente, se abocaron a los mecanismos neurales centrales de la emoción por la acción combinada de la rama simpática y parasimpática del sistema nervioso autónomo y consideraron que los cambios fisiológicos eran similares para las distintas emociones, con una variación solamente en la intensidad con que se expresaban (Cannon, 1927a, 1927b).

El grado de activación de las variables autonómicas vinculadas con la ansiedad se ha inferido a partir de los valores en la actividad eléctrica de la piel: la conductividad de la piel, y en la actividad cardiovascular: la tasa cardíaca y la presión sanguínea. Las variables somáticas se han infe-

rido a través de la actividad electromiográfica de los músculos: la actividad eléctrica de los músculos del esqueleto (Carrobbles & Buela-Casal, 1997; Vila, 1998, 2004). Las respuestas fisiológicas más utilizadas para estudiar los mecanismos psicológicos del miedo y la ansiedad proceden del estudio de los reflejos protectores. Estos evitan daños y lesiones, actúan como inhibidores conductuales y facilitan la detección de estímulos amenazantes. Principalmente se han valorado el reflejo de sobresalto y el reflejo cardíaco de defensa. El primero es un patrón de activación motora que se produce por estimulación intensa o aversiva de inicio repentino. Es una reacción refleja, innata, involuntaria y sujeta a condicionamiento con tres componentes principales: parpadeo, inclinación brusca de la cabeza hacia adelante y una onda de flexión que se trasmite desde el tronco hacia las rodillas (Landis & Hunt, 1939). Se mide a partir del registro psicofisiológico del parpadeo (Lang, Bradley, & Cuthbert, 1997). El segundo alude al componente cardíaco de la respuesta ante estimulación discreta intensa o aversiva. Se mide a través del cambio en la tasa cardíaca (latidos por minuto) durante los 80 segundos posteriores a la presentación del estímulo (Vila, 1998).

El registro de las variables psicofisiológicas requiere de una serie de fases: (a) detección de la señal proveniente del organismo; (b) transformación de las señales del organismo a señales eléctricas; (c) amplificación de dichas señales; (d) registro propiamente dicho: gráfico, digital y/o computarizado y (e) conversión de la señal registrada para su análisis estadístico (Calero et al., 2013; Carrobbles & Buela-Casal, 1997). Dicho procedimiento ha reportado diversas fuentes de error procedentes de las condiciones físicas del ambiente, la calidad del instrumental, el nivel previo de activación, la discriminación de respuestas no relacionadas con la estimulación, la fluctuación temporal de las variables fisiológicas

y los patrones atribuibles a diferencias personales no generalizables, entre otras. Asimismo, la artificialidad de la situación dificulta la inducción de estados emocionales y limita la transferencia y generalización de los resultados. No obstante, actualmente existen medidores portables que permiten la evaluación de las variables fisiológicas en situaciones naturales, lo que ha posibilitado su aplicación en numerosos campos: psicología clínica, medicina conductual, neuropsicología, ergonomía y deporte, entre otros (Calero et al., 2013; Carrobbles & Buela-Casal, 1997).

Evaluación indirecta de las variables fisiológicas

La evaluación indirecta surgió como una opción válida cuando no se contaba con el equipamiento técnico o profesional requerido o cuando por cuestiones de diseño se optaba por modelos no experimentales o clínicos (Carrobbles & Buela-Casal, 1997). Fundamentalmente, se ha realizado a través de técnicas de observación y técnicas de autoinforme. Estas últimas pueden clasificarse según Fernández-Ballesteros (2013c) en: autorregistros, escalas/inventarios/cuestionarios y entrevistas.

La evaluación a través de la observación es posible sólo cuando existen manifestaciones externas fácilmente discriminables. Supone un procedimiento sistemático de recolección de información a partir de la percepción deliberada, por parte del observador entrenado, de ciertas manifestaciones emitidas por un sujeto (Fernández-Ballesteros, 2013b). Las unidades de análisis en este caso han sido las manifestaciones externas de las variables psicofisiológicas definidas previamente.

El autorregistro se ha utilizado principalmente para el registro de los síntomas asociados a determinados estados fisiológicos de fácil detección para el sujeto. No se ha utilizado para la

evaluación del estado fisiológico en sí mismo porque este no es susceptible de evaluación a través de este procedimiento. Su detección en variables como tasa cardíaca, actividad electrodérmica y ondas cerebrales ha podido mejorarse con el entrenamiento y con la utilización de métodos de registro adecuados y/o portables en situaciones naturales (Buela-Casal & Sierra, 2004; Carrobles & Buela-Casal, 1997).

Las escalas, cuestionarios o inventarios han sido utilizados para evaluar de forma específica o general una serie de manifestaciones o síntomas asociados a los estados psicofisiológicos de las emociones, tales como el miedo y la ansiedad. Requieren, al igual que el autorregistro, que el sujeto posea la capacidad de discriminación y detección de una manifestación propia. Los instrumentos específicos más utilizados han sido: (a) Cuestionario sobre Percepción Autónoma (APQ; Mandler, Mandler, & Ulville, 1958), (b) Inventario de Indicadores Psicofisiológicos (Cautela & Upper, 1976) y (c) Inventario de Ansiedad E-R (Endler, Hunt, & Rosenstein, 1962). Por su parte, los instrumentos generales para evaluar síntomas clínicos de ansiedad, que incluyen síntomas psicofisiológicos han sido: (a) Escala de Hamilton (HARS; Hamilton, 1959); (b) Inventario de Ansiedad de Beck (BAI; Beck, Epstein, Brown, & Steer, 1988); (c) Inventario Revisado de 90 síntomas (SCL-R-90; Derogatis, 1977); (d) Inventario de Situaciones y Respuestas de Ansiedad (ISRA; Miguel-Tobal & Cano-Vindel, 1986); (e) Escala de Ansiedad de Zung (EAZ; Zung, 1971); (f) Escala de Ansiedad y Depresión Hospitalaria (HADS; Zigmond & Snaith, 1983); (g) Escala de Ansiedad Clínica (CAS; Snaith, Baugh, Clayden, Husain, & Sipple, 1982); (h) Escala Breve de Ansiedad de Tyrer (BSA; Tyrer, Owen, & Cicchetti, 1984) e (i) Cuestionario de Detección de Ansiedad (ASQ-15; Wittchen & Boyer, 1998), entre otros.

Finalmente, la entrevista también ha permitido relevar la presencia y magnitud de los síntomas asociados a los estados psicofisiológicos de la ansiedad. Construidos para aumentar la confiabilidad y validez diagnóstica a través de la estandarización del proceso de evaluación, los criterios internacionales y la indagación sistemática y exhaustiva de síntomas han permitido la inclusión confiable de los pacientes dentro de las categorías diagnósticas y el diagnóstico diferencial en los cuadros limítrofes (Echeburúa, 1996). Específicamente para la evaluación de los trastornos de ansiedad, las entrevistas más utilizadas han sido: (a) Entrevista para los Trastornos de Ansiedad según el DSM-IV (ADIS-IV; Brown, Di Nardo, & Barlow, 1994), la cual es una entrevista semiestructurada heteroaplicada con la que se realiza una evaluación de los trastornos de ansiedad y sus trastornos comórbidos más comunes; (b) Entrevista Clínica Estructurada para el DSM-IV y DSM-5 (SCID-IV y SCID-5; First, Spitzer, Gibbon, Hilsenroth, & Segal, 2004, 2015), la cual es una entrevista semiestructurada destinada a realizar los diagnósticos más importantes del eje I del DSM-IV y los diagnósticos equivalentes del DSM-5; (c) Mini Entrevista Neuropsiquiátrica Internacional (MINI; Lecrubier et al., 1997), la cual es una entrevista breve pero de gran difusión que explora los principales trastornos psiquiátricos del eje I del DSM-IV y del CIE-10.

Evaluación de la dimensión conductual

La *conducta* es una de las variables más estudiadas y evaluadas en la historia de la psicología. Cada modelo teórico conceptualizó y abordó la conducta desde una perspectiva distinta, diversificándose las estrategias y técnicas utilizadas para su evaluación. Por ejemplo, en términos generales, el modelo psicoanalítico consideró la

conducta como un signo que necesitaba ser interpretado simbólicamente para adquirir un valor diagnóstico. Las conductas manifiestas, actos fallidos, sueños o ejecución en técnicas proyectivas fueron considerados como el resultado de factores intrapsíquicos, principal objeto de estudio de dicho modelo. Las técnicas proyectivas han sido instrumentos sensibles para revelar aspectos inconscientes de la conducta a partir de una variedad de respuestas subjetivas (Calero et al., 2013). Por su parte, para el modelo psicométrico de rasgos, la conducta también estaba determinada por los atributos intrapsíquicos (llamados rasgos, habilidades o intereses). Al igual que en el modelo psicodinámico, dichas variables no podían evaluarse de forma directa, sino a través de la conducta manifiesta. Esta evaluación requería, en este caso, de una interpretación estadística. Utilizaron tests estandarizados construidos según estrategias teórico racionales, empíricas y factoriales, apoyadas en el supuesto de generalidad o estabilidad de las conductas (Buela-Casal & Sierra, 1997; Calero & Padilla, 2013; Fernández-Ballesteros, 2013a).

Por el contrario, el modelo conductual planteó la especificidad situacional de la conducta y, por ende, su dependencia de las variables exógenas o ambientales. Consideró la evaluación como un componente del propio proceso terapéutico, inseparable de la modificación conductual (Buela-Casal & Sierra, 1997; Fernández-Ballesteros, 1994, 2013a). Incluyó distintos enfoques de evaluación producto de su desarrollo conceptual, metodológico y de la diversificación de sus variables de estudio. Por ejemplo, el modelo clásico tomó como objeto de estudio las respuestas motoras y los estímulos físicos subyacentes y el modelo mediacional incorporó la aplicación de los procesos de aprendizaje abocándose al análisis aplicado de la conducta en función de sus variables situacionales. Consideró la evaluación y la intervención como parte de un mismo proceso destinado a la

modificación de la conducta. Si bien no incluyó los procesos fisiológicos ni cognitivos como objeto de estudio, tampoco negó su existencia considerándolos productos de la conducta. Por su parte, el modelo conductual interactivo integró las relaciones entre las variables fisiológicas, cognitivas, situacionales y conductuales, a las que incluyó los estados afectivos o emocionales. Consideró que la conducta podía ser causada o activada por dichos factores aislados o combinados (Buela-Casal & Sierra, 1997; Fernández-Ballesteros, 2013a).

La evaluación conductual se ha definido como la identificación de unidades de respuesta significativas y sus variables generadoras con el fin de comprender y modificar la conducta (Hayes, Nelson, & Jarrett, 1986). Supone que el comportamiento puede ser funcionalmente explicado a través de ciertas condiciones ambientales actuales que controlan o mantienen dichas conductas y por ciertas variables personales constituidas a partir de condiciones ambientales pasadas. Estas últimas, en interacción con el organismo, conformarían la personalidad del sujeto. Es un proceso que implica la aplicación de distintos procedimientos de evaluación a través de una serie de fases que requieren cada vez de instrumentos más estrictos, sofisticados y precisos (Fernández-Ballesteros, 1994). A pesar de su diversidad, a grandes rasgos pueden clasificarse en métodos directos y métodos indirectos (Barrios, 1988). Los primeros aluden a los métodos conductuales clásicos que evalúan la conducta en el momento y en el lugar de su ocurrencia real. Los segundos, a métodos que lo hacen en otro tiempo y/o lugar.

Evaluación directa de las variables conductuales

La evaluación directa de las variables conductuales se ha realizado principalmente a través del registro de la actividad motora y la observa-

ción sistemática. La actividad motora se evalúa a través del registro instrumental del movimiento en el espacio y la aceleración, lo que ha permitido obtener muestras de la actividad motora naturales, objetivas, confiables, unidimensionales, estables y válidas en condiciones de laboratorio. Sin embargo, requiere equipamiento y profesionales capacitados (Tryon, 1997). A pesar de la difusión que tuvieron en su momento, su uso en la práctica clínica fue limitado, ya que se preferían técnicas creadas específicamente para el registro de la conducta, como la observación o los autoinformes. Sin embargo, con el surgimiento de medidores portables, en la última década se ha podido retomar la evaluación objetiva de las variables fisiológicas en situaciones naturales (Buela-Casal & Sierra, 2004; Calero et al., 2013).

La observación sistemática ha sido la técnica principal y más eficaz para registrar las conductas motoras de ansiedad. Permite cuantificar la conducta espontánea del sujeto en situaciones naturales en el momento de su ocurrencia. Implica la percepción deliberada, el registro, la codificación y el análisis de dicha conducta a partir de un procedimiento sistematizado que permite la replicabilidad y el control de los resultados. Su objetivo principal es registrar medidas de conducta que posean variabilidad en un intervalo de tiempo determinado. Las unidades de medida más utilizadas han sido: ocurrencia, orden, frecuencia y duración. Requiere de un proceso riguroso de codificación y categorización que posibilite la sistematización de los datos recabados y garantice la precisión de los resultados obtenidos. La incorporación de dispositivos portátiles de registro ha facilitado dicha tarea (Buela-Casal & Sierra, 2004). Sin embargo, esta técnica ha reportado diversas fuentes de error procedentes de la reactividad del sujeto, del grado de participación del observador, sus expectativas y entrenamiento, y del sistema de observación (tipo de registro elegido, procedimientos y dispo-

sitivos técnicos; Fernández-Ballesteros, 2013b; Quera & Behar, 1997).

Evaluación indirecta de las variables conductuales

La evaluación indirecta de la conducta se fue adoptando de forma progresiva en la práctica clínica y de investigación conductual. En las últimas décadas los autoinformes (entrevistas, autorregistros, escalas, cuestionarios e inventarios) han sido las técnicas de evaluación más utilizadas (Buela-Casal & Sierra, 2004; Fernández-Ballesteros, 1994, 2013c; Jensen, 1996). En este contexto, han adquirido ciertas características distintivas al ser considerados como una muestra de un tipo determinado de conducta y estar relacionados con situaciones específicas (Del Barrio, 1998; Fernández-Ballesteros, 2013c).

En la práctica clínica se han utilizado dos modelos de entrevista: diagnóstica y conductual. La primera busca la exploración diagnóstica y psicopatológica de los trastornos clínicos de acuerdo con los criterios internacionales (Sierra, Buela-Casal, & Fernández, 2004). Remite a una información general sobre un trastorno, su etiología, desarrollo típico, pronóstico y tratamiento existentes. Dicha información se generaliza a través de diferencias individuales, lo que orienta la práctica e investigación clínica (Vizcarro, 1998). La evaluación clínica requiere, adicionalmente, una evaluación ideográfica más o menos detallada en la búsqueda de leyes particulares que complementen la información general (Fernández-Ballesteros, 2013a; Silva, 1998). Las entrevistas estandarizadas ya citadas incluyen la evaluación del comportamiento y sus alteraciones, información crucial para determinar la adaptación y el deterioro funcional del sujeto en áreas importantes de su vida, requisito necesario para diagnosticar

un trastorno mental (*American Psychiatric Association*, 2013). Por su parte, la entrevista conductual es una herramienta esencial e indispensable del proceso de evaluación y tratamiento. Pretende relevar información detallada y precisa sobre las conductas problema, los contextos de desarrollo y las consecuencias funcionales y adaptativas para el sujeto. Permite establecer prioridades de intervención y proporciona una dirección operacionada para la intervención terapéutica destinada a la modificación de la conducta problema (*Buela-Casal & Sierra*, 2004; *Bruch & Meyer*, 1996). En la actualidad, tal como en décadas anteriores, la entrevista continúa siendo la técnica más utilizada en la práctica clínica como instrumento de evaluación e intervención psicológica (*Márquez*, 2013; *Sierra et al.*, 2004).

El autorregistro implica un procedimiento combinado de la técnica de observación y el registro de ciertos aspectos del propio comportamiento como son: ocurrencia, duración, intensidad, antecedentes, consecuentes y resultado. Aporta datos a la evaluación conductual y promueve el cambio (*Fernández-Ballesteros*, 2013b, 2013c). Es el único método que permite la evaluación de las interacciones en la vida diaria del paciente, la detección de los elementos situacionales y las respuestas más significativas. Sin embargo, esta técnica ha planteado limitaciones en individuos defensivos o con baja capacidad de introspección (*Buela-Casal & Sierra*, 2004; *Echeburúa*, 1996).

Por otro lado, las escalas, cuestionarios e inventarios han cobrado un papel importante en la evaluación conductual actual y son utilizados como instrumentos primordiales para precisar la información recogida con técnicas de amplio espectro como la entrevista, la observación sistemática, la autobiografía, entre otras. Son instrumentos complementarios que, si bien ofrecen escasa especificidad situacional y de respuesta (*Buela-Casal & Sierra*, 2004), permiten compa-

rar los comportamientos del sujeto con su grupo normativo (*Fernández-Ballesteros*, 1994, 2013b; *Jensen*, 1996). Particularmente en la evaluación clínica de los trastornos de ansiedad, el autoinforme ha sido la técnica más utilizada por los evaluadores y terapeutas conductuales (*Buela-Casal & Sierra*, 2004; *Menéndez-Carrillo & Maciá-Antón*, 1994). A pesar de su difusión, son pocas las escalas que incluyen enunciados con manifestaciones conductuales de ansiedad. De las antes mencionadas, solo el ISRA (*Miguel-Tobal & Cano-Vindel*, 1986), incorpora dicha dimensión, a la que denomina respuestas motoras.

Evaluación de la dimensión cognitiva

Las variables *cognitivas* de la ansiedad comenzaron a evaluarse a partir del desarrollo de la teoría cognitivo-conductual. Esta consideró a los procesos cognitivos que mediaban entre un estímulo aversivo y la respuesta emocional del sujeto como responsables del desarrollo y mantenimiento de la ansiedad. La valoración cognitiva de sujetos con trastornos de ansiedad permitió comprender su etiología y persistencia (*Clark & Beck*, 2012), planificar el tratamiento clínico y asegurar su éxito (*Mohlman & Gorman*, 2005). Junto con la formulación cognitiva del caso, se transformaron en pilares primordiales para una psicoterapia efectiva, sirviendo de nexo entre la teoría cognitivo-conductual a y el tratamiento (*Clark & Beck*, 2012).

Las variables cognitivas operacionales y de contenido han sido el principal objeto de estudio para este modelo (*Buela-Casal & Sierra*, 1997; *Fernández-Ballesteros*, 2013a). Las primeras aluden a los mecanismos mediante los cuales el sistema opera procesando la información: memoria, atención, codificación, entre otros. Las segundas refieren a los contenidos de la información y sus

representaciones en proposiciones, pensamientos, imágenes, esquemas, etc. El grado de accesibilidad de lo cognitivo es lo que ha determinado el método idóneo para acceder a dichas variables (Pérez-Pareja, 1997); en la evaluación de los contenidos cognitivos se han preferido los autoinformes, y en la evaluación de las operaciones cognitivas, las técnicas objetivas.

Evaluación de los contenidos cognitivos

Según Fernández-Ballesteros (2013c), el autoinforme (entrevista, autorregistro, escala/cuestionario/inventario) es el método prioritario y directo por excelencia para explorar los contenidos cognitivos. Es la única vía de acceso a las manifestaciones internas del sujeto, expresadas a través de un mensaje verbal, producto de la introspección o la autoobservación. Los contenidos y productos cognitivos son accesibles a nivel de la conciencia e informables a través de dicha verbalización. La entrevista ha sido la técnica de autoinforme más utilizada para evaluar los contenidos cognitivos (Fernández-Ballesteros, 2013c; Silva, 1998). Las tres entrevistas estandarizadas antes mencionadas para valorar la presencia y severidad de los síntomas típicos de los trastornos de ansiedad de acuerdo con los criterios diagnósticos internacionales también han sido utilizadas para evaluar la dimensión cognitiva, sobre todo a nivel de pensamientos automáticos, imágenes intrusivas e inferencias. Sin embargo, su utilización suele ser engorrosa, demanda mucho tiempo de aplicación y una sólida formación clínica del entrevistador (Buela-Casal & Sierra, 2004).

Como alternativa a esta, el autorregistro ha sido la segunda alternativa para la evaluación de las cogniciones y el sesgo cognitivo a nivel de pensamientos, especialmente en la clínica e investigación bajo el modelo cognitivo. Su sencillez y

aplicabilidad en situaciones naturales han favorecido su utilización. Supone, por un lado, atender de forma deliberada a las cogniciones en el momento en que estas se producen (automonitoreo) y, por otro, su registro mediante un procedimiento preestablecido (Buela-Casal & Sierra, 2004; Fernández-Ballesteros, 2013c). Esto le permite al paciente instaurar el procesamiento cognitivo reflexivo como respuesta mediadora a situaciones donde primaba el procesamiento cognitivo automático (Clark & Beck, 2012). Al admitir el registro de varios aspectos a la vez permite en la clínica cognitiva relevar y evaluar variables de desarrollo y mantenimiento de la ansiedad “en tiempo real” (Clark & Beck, 2012, p. 151). Los autorregistros más utilizados en este ámbito han sido: (a) registros de pensamientos automáticos, (b) valoración diaria de ansiedad y registro de situaciones, (c) registro de pensamientos aprensivos para automonitoreo y (d) registro de valoración, entre otros. Los mismos fueron incluidos por Clark y Beck (2012) en los protocolos de evaluación y tratamiento que validaron. No obstante, existen algunos elementos que han afectado la validez de este procedimiento, por ejemplo, la motivación, la reactividad, la falta de conciencia de las conductas implicadas, la complejidad de la interacción de las respuestas, su grado de accesibilidad, las imprecisiones en el recuerdo o en la comprensión de la situación, la complejidad o inadecuación del formato de registro (Buela-Casal & Sierra, 2004; Echeburúa, 1996).

Las escalas, cuestionarios e inventarios han sido la tercera modalidad utilizada para evaluar los contenidos cognitivos y las distorsiones cognitivas. Las más difundidas son: (a) Listado de Comprobación de Cogniciones (CCL; Beck, Brown, Steer, Eidelson, & Riskind, 1987), incluye una subescala de ansiedad (CCL-A) que evalúa la frecuencia de pensamientos ansiosos negativos en torno a temas de incertidumbre y orientación

hacia el futuro; (b) Cuestionario de Estados de Preocupación Penn (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990), evalúa la propensión a la preocupación y la intensidad con que se presentan las experiencias de preocupación, sin hacer referencia a temas específicos; (c) Cuestionario de Pensamientos Automáticos (ATQ; Hollon & Kendall, 1980), evalúa pensamientos irracionales asociados a la depresión y la ansiedad y (d) Cuestionario de Esquemas Maladaptativos Tempranos (YSQ L2; Young, 1999), evalúa la presencia de dieciséis esquemas disfuncionales desarrollados durante la infancia, estables y duraderos a lo largo de la vida de un individuo, que sirven como marcos para el procesamiento de experiencias posteriores. Asimismo, dentro de las escalas antes mencionadas para evaluar síntomas clínicos generales de ansiedad, las siguientes incluyen información sobre el dominio cognitivo: (a) Inventario de Situaciones y Respuestas de Ansiedad (ISRA; Miguel-Tobal & Cano-Vindel, 1986); (b) Inventario de Ansiedad Estado Rasgo (STAI; Spielberger, Gorsuch, & Lushene, 1983); (c) Listado Revisado de 90 síntomas (SCL-R-90; Derogatis, 1977); (d) Inventario de Ansiedad de Beck (BAI; Beck et al., 1988); (e) Escala de Ansiedad de Zung (EAZ; Zung, 1971); (f) Escala de Ansiedad de Hamilton Autoaplicada (HARS; Hamilton, 1959); (g) Escala de Ansiedad y Depresión Hospitalaria (HADS; Zigmond & Snaith, 1983); (h) Escala de Ansiedad Clínica (CAS; Snaith et al., 1982); (i) Escala Breve de Ansiedad de Tyrer (BSA; Tyrer et al., 1984) y (j) Cuestionario de Detección de Ansiedad (ASQ-15; Wittchen & Boyer, 1998).

Evaluación de las operaciones cognitivas

Implican procesos secuenciales de procesamiento de la información que no son transmisibles verbalmente por el sujeto, principalmente

porque se producen de forma automática y no son accesibles a la conciencia, y secundariamente porque son muy reactivos a la introspección que implica el autoinforme. Por ello, para su evaluación se ha preferido la utilización de técnicas objetivas, como el tiempo de reacción y la respuesta a serie de palabras, entre otras, que admiten la manipulación de los estímulos y el registro automático de las respuestas en base al número de aciertos y errores, según ciertos parámetros temporales (Calero et al., 2013). Las operaciones cognitivas más evaluadas en los trastornos de ansiedad han sido la codificación y recuperación de la información, la velocidad de procesamiento, la atención selectiva, la atención autofocalizada y la propagación de la activación (Clark & Beck, 2012). La evaluación del funcionamiento atencional ha ocupado un lugar preponderante; los mecanismos considerados para explicar los sesgos atencionales han sido múltiples, implementados por diferentes operaciones cognitivas (Eysenck, 1988; Eysenck, Derakshan, Santos, & Calvo, 2007). La tarea más utilizada en investigación experimental para evaluar los sesgos atencionales en sujetos ansiosos ha sido la Stroop emocional, que permitió detectar una lentificación de las respuestas al nombrar el color de palabras específicas ansiógenas, en comparación con las palabras neutras (Mathews & MacLeod, 1985). Sin embargo, dicho hallazgo presenta limitaciones interpretativas, no se puede asegurar fehacientemente que la denominación más lenta de la palabra sea producto de un efecto de interferencia por el significado que posee (Clark & Beck, 2012) ni consecuencia de un sesgo atencional selectivo hacia los contenidos amenazantes (Pacheco-Unguetti, Lupiáñez, & Acosta, 2009).

Evaluación de la dimensión afectiva

La dimensión *afectiva* comenzó a considerarse y valorarse con el desarrollo de la teoría cognitivo-conductual. Para Clark y Beck (2012), esta derivaba de la activación conjunta de las dimensiones *fisiológica* y *cognitiva* y constituía la experiencia subjetiva de la sensación ansiosa, emergente de la valoración cognitiva que el sujeto realizaba principalmente de sus manifestaciones de ansiedad. A pesar de que algunos percibían como perturbadores o amenazantes los síntomas *fisiológicos*, otros, los fenómenos *cognitivos* (preocupación o pensamientos intrusos indeseados), y otros, el aumento de la sensación generalizada de ansiedad, la experiencia subjetiva de sentir ansiedad era lo que la persona consideraba como amenazante e intolerable y a la cual reaccionaba de forma defensiva a través de las conductas evitativas (Clark & Beck, 2012).

Para la teoría cognitivo-conductual, la experiencia subjetiva de ansiedad es la que motiva a las personas a la consulta clínica. Su eliminación ha sido el principal criterio para determinar el éxito del tratamiento, cuyo primer objetivo de intervención se propone normalizar la experiencia de ansiedad. Dado que el paciente se ve compelido a consultar por la perturbación que le produce dicha experiencia ansiosa, evaluarla y abordarla permite aumentar la motivación para el cambio y propiciar la alianza terapéutica. Según Clark y Beck (2012), los pacientes solían mostrarse más dispuestos a aceptar el tratamiento si sus experiencias ansiosas eran abordadas con inmediata relevancia. Por ende, la evaluación de la dimensión afectiva resulta primordial para el diagnóstico, la alianza terapéutica y la eficacia del tratamiento en los trastornos de ansiedad.

Como se mencionó con anterioridad, para Fernández-Ballesteros (2013c) el autoinforme es el procedimiento más directo y prioritario para

evaluar los contenidos mentales. Dentro de este, las entrevistas estandarizadas ya citadas incluyen pautas para la valoración de las manifestaciones afectivas, por lo que también han sido utilizadas para evaluar esta dimensión. Por su parte, los autorregistros permiten evaluar la experiencia subjetiva de ansiedad a partir del registro de las situaciones, cogniciones y respuestas de ansiedad que convergen simultáneamente en una situación de interacción. Esta información facilita determinar si la experiencia ansiosa es excesiva o desproporcionada teniendo en cuenta los factores situacionales, culturales y evolutivos. Asimismo, se ha extendido el uso de escalas para evaluar el componente afectivo de los trastornos emocionales. A partir del modelo bidimensional del afecto propuesto por Watson y Clark (1984) sobre la estructura básica del afecto, se desarrolló una escala específica para evaluar la afectividad: la Escala de Afecto Positivo y Negativo (PANAS; Watson, Clark, & Tellegen, 1988). Watson y Clark (1984) diferenciaron dos dimensiones relativamente independientes del afecto, denominadas afecto positivo y afecto negativo. El *afecto positivo* representa la dimensión de emocionalidad placentera, caracterizada por la motivación, energía, concentración, alerta y participación agradable. En contraste, el *afecto negativo* representa la dimensión de emocionalidad displacentera, caracterizada por el malestar, el desinterés subjetivo y la participación desagradable e incluye una variedad de estados emocionales aversivos como el disgusto, la ira, el miedo, la culpa y el nerviosismo (Watson et al., 1988). Este modelo planteó importantes implicaciones en los trastornos de ansiedad y del estado de ánimo al describir características afectivas que son propias de cada cuadro. En la depresión hallaron baja afectividad positiva y alta afectividad negativa, mientras que en la ansiedad, elevada hiperactivación fisiológica y alta afectividad negativa. Por ende, ambas poseen un componente

común (el afecto negativo), y componentes específicos (bajo afecto positivo y alta hiperactivación fisiológica, respectivamente).

Aplicaciones clínicas de la evaluación multidimensional de la ansiedad

La evaluación dimensional de la ansiedad proporciona un marco para comprender cómo se manifiesta un trastorno en un paciente determinado y los factores que mantienen la conducta disfuncional. Dicha información resulta beneficiosa para la conceptualización del caso, guía al terapeuta en la conducción de tratamientos eficaces al permitir la planificación de intervenciones efectivas (Beck, 2000; Clark & Beck, 2012; Keegan & Holas, 2010). De acuerdo con la evidencia empírica, se emplea un tipo de intervención cuando predominan los síntomas *cognitivos* y *afectivos* y otro distinto cuando predominan los síntomas *fisiológicos* o *conductuales*. Por ejemplo, las investigaciones de Lang, Melamed y Hart (1970) y Wolpe (1977) encontraron que la técnica de desensibilización sistemática era más efectiva para tratar los síntomas *fisiológicos* que los *cognitivos*. Por su parte, otros autores (Öst, Jerremalm, & Johansson, 1981; Öst, Johansson & Jerremalm, 1982) encontraron que aquellos que presentaban mayor reactividad fisiológica se beneficiaban con técnicas de relajación, mientras que aquellos con mayor reactividad motora lo hacían con técnicas de entrenamiento en habilidades sociales. Clark y Beck (2012), a partir de la evidencia acumulada, clasificaron las intervenciones en *cognitivas* y *conductuales*. Las primeras fueron utilizadas tanto para tratar los síntomas cognitivos como afectivos, incrementan la flexibilidad cognitiva a través del tratamiento de los pensamientos, valoraciones y creencias ansiosas. Poseen como objetivo primordial cambiar el foco de la amenaza, modificar

la amenaza, la vulnerabilidad y las creencias sesgadas, normalizar el miedo y la ansiedad, fortalecer la eficacia personal y resolver los problemas relacionados con la búsqueda de seguridad. Por su parte, las intervenciones *conductuales* se han utilizado para modificar los patrones de respuesta conductual disfuncionales del paciente, así como también para aliviar y tratar pacientes con gran activación de respuestas fisiológicas. Sin embargo, cabe acotar que para la teoría cognitivo conductual, la cognición, la emoción y la conducta son subsistemas interdependientes. La modificación de un subsistema produce indefectiblemente cambios en los otros dos. Asimismo, un cambio en la respuesta conductual o fisiológica requiere de la intervención sobre los sesgos cognitivos asociados que las generan o sostienen.

La eficacia terapéutica se establece a través del éxito del plan de tratamiento, o sea, del logro de los objetivos propuestos y del efecto esperado (Castro-Solano, 2003; Gómez-Penedo & Roussos, 2012; Bados-López, García-Grau, & Fusté-Escolano, 2002). Si bien no hay acuerdo sobre las variables a evaluar ni sobre los criterios de éxito, la medición a través de instrumentos estandarizados resulta básica para determinar la significación estadística o clínica y el tamaño del efecto de una intervención o tratamiento (Gómez-Penedo & Roussos, 2012). Las mejorías más evidentes son las que se producen sobre los cambios sintomáticos. La utilización de medidas estandarizadas permite comparar el puntaje que obtiene un paciente con la población a la que pertenece y determinar el efecto de la intervención sobre el perfil sintomático, comparando su puntuación en diferentes momentos del proceso terapéutico. Usualmente, los clínicos se valen de entrevistas diagnósticas estructuradas y de diversas medidas estandarizadas para valorar la frecuencia y severidad sintomatológica. Según Clark y Beck (2012), estas últimas son de gran utilidad porque

permiten detectar los síntomas más relevantes, brindan una medida de su gravedad, permiten su comparación con la población normativa y posibilitan su administración repetida a lo largo del curso del tratamiento.

Alcances de la evaluación multidimensional en investigación clínica

Clásicamente, cada sistema de respuesta se ha evaluado por separado, aplicándose en cada caso el método más idóneo (Fernández-Ballesteros, 1994). Sin embargo, esta metodología presentó diversas limitaciones que no justifican su evaluación por separado. Por un lado, aumentó la varianza intermétodo que hacía aumentar artificialmente la discordancia y desincronía entre sistemas de respuesta (Cano-Vindel & Miguel-Tobal, 1990; Miguel-Tobal, 1990); por otro, no aportó información independiente que resulte significativa debido a la elevada correlación entre los instrumentos (Echeburúa, 1996). Es por ello que generalmente se opta por la aplicación de métodos menos costosos y fáciles de utilizar.

Tanto en la práctica clínica como en investigación, el autoinforme en su formato de escalas, cuestionarios e inventarios ha desplazado a los métodos conductuales clásicos (entrevista conductual y observación sistemática) y a los registros psicofisiológicos como instrumentos primordiales para la evaluación de los trastornos de ansiedad (Buela-Casal & Sierra, 2004; Echeburúa, 1996; Fernández-Ballesteros, 1994; 2013c). Por ende, no solo han demostrado ser útiles para evaluar el sistema cognitivo, sino que también permiten obtener una primera aproximación a los sistemas fisiológicos y motores, lo que la impone como la técnica más empleada para determinar la intensidad sintomática y cuantificar los cambios terapéuticos (Buela-Casal & Sierra, 2004; Cano-Vindel

& Miguel-Tobal, 1990; Miguel-Tobal, 1990; Fernández-Ballesteros, 1994).

Por otro lado, si bien a partir de la década de 1960 se reconoció la necesidad de una evaluación conjunta de diversos sistemas de respuestas, a la hora de abordar el fenómeno de la ansiedad aún son escasas las investigaciones clínicas que la incluyen. Una revisión sistemática realizada por Cano-Vindel y Miguel-Tobal (1990) sobre la forma y los sistemas de evaluación en investigación clínica concluyó que, hasta entonces, existía un predominio de la evaluación de la dimensión cognitiva y motora sobre la fisiológica, principalmente a través de técnicas de autoinforme.

Conclusión

Este trabajo proporciona una descripción instrumental sobre la evaluación de la ansiedad y su evolución desde los primeros modelos psicofisiológicos de las emociones. Realiza un recorrido por las diversas técnicas y propuestas instrumentales adoptadas por los distintos modelos teóricos para evaluar cada dimensión. Repasa las aplicaciones clínicas y los alcances en investigación que posee la evaluación multidimensional, planteando la necesidad de una evaluación que logre abarcar la complejidad del constructo a fin de aumentar la eficacia de los tratamientos.

Según lo expuesto, existe un amplio repertorio instrumental para la evaluación multidimensional de la ansiedad con métodos específicos para medir cada una de sus dimensiones. Sin embargo, la valoración multimétodo ha planteado diversas limitaciones prácticas y metodológicas. Si bien las técnicas de autoinforme parecen ser la respuesta tecnológica más versátil y económica para obtener una primera aproximación a las dimensiones fisiológicas y conductuales, aún son escasos los instrumentos que incorporan dichas dimensio-

nes, lo que crea una vacante en este aspecto. Es por ello que a la hora de evaluar la ansiedad se hace imperioso utilizar el método que permita la evaluación más completa posible del constructo teniendo en cuenta su multidimensionalidad.

Referencias

- American Psychiatric Association (2013). *DSM-5. Manual diagnóstico y estadístico de los trastornos mentales* (5ª ed.). Madrid, España: Médica Panamericana.
- Bados-López, A., García-Grau, E., & Fusté-Escolano, A. (2002). Eficacia y utilidad clínica de la terapia psicológica. *International Journal of Clinical and Health Psychology*, 2(3), 477-502. Recuperado de <http://www.aepc.es/ijchp/busca.php?coid=Espa%F1ol>
- Barrios, B. A. (1988). On the changing nature of behavioral assessment. En A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3ª ed., pp. 3-41). New York, NY: Pergamon.
- Beck, A. T., Brown, G., Steer, R. A., Eidelson, J. I., & Riskind, J. H. (1987). Differentiating anxiety and depression: A test of the cognitive content-specificity hypothesis. *Journal of Abnormal Psychology*, 96(3), 179-183. doi: [10.1037//0021-843x.96.3.179](https://doi.org/10.1037//0021-843x.96.3.179)
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56(6), 893-897. doi: [10.1037//0022-006x.56.6.893](https://doi.org/10.1037//0022-006x.56.6.893)
- Beck, J. S. (2000). *Terapia cognitiva: Conceptos básicos y profundización*. Barcelona, España: Gedisa.
- Brown, T. A., Di Nardo, P. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview for DSM-IV (ADIS-IV)*. San Antonio, TX: The Psychological Corporation.
- Bruch, M., & Meyer, V. (1996). La entrevista conductual. En G. B. Buela-Casal, V. E. Caballo & J. C. Sierra (Eds.), *Manual de evaluación en psicología clínica y de la salud* (pp. 85-107). Madrid, España: Siglo XXI.
- Buela-Casal, G., & Sierra, J. C. (1997). Modelos de evaluación psicológica. En G. Buela-Casal & J. C. Sierra (Eds.), *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones* (pp. 103-145). Madrid, España: Siglo XXI.
- Buela-Casal, G., & Sierra, J. C. (2004). *Manual de evaluación y tratamientos psicológicos*. Madrid, España: Biblioteca Nueva.
- Calero, M. D., & Padilla, J. L. (2013). Técnicas psicométricas: Los test. En R. Fernández-Ballesteros (Ed.), *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp. 299-337). Madrid, España: Pirámide.
- Calero, M. D., Márquez, M. O., Vizcarro, C., & Fernández-Ballesteros, R. (2013). Otras técnicas: Objetivas y proyectivas. En R. Fernández-Ballesteros (Ed.), *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp. 275-298). Madrid, España: Pirámide.
- Cannon, W. B. (1927a). The James-Lange's theory of emotion: A critical examination and an alteration. *The American Journal of Psychology*, 39(1-4), 106-124. doi: [10.2307/1415404](https://doi.org/10.2307/1415404)
- Cannon, W. B. (1927b). *Bodily changes in pain, hunger, fear and rage*. New York and London: Appleton and Company.
- Cano-Vindel, A. C., & Miguel-Tobal, J. J. (1990). Revisión de la evaluación de los tres sistemas de respuesta de ansiedad en las terapias cognitivo-conductuales. *Terapia del Comportamiento*, 44-45, 109-114. Recuperado de: https://www.researchgate.net/publication/230577069_Revision_de_la_evaluacion_de_los_tres_sistemas_de_respuesta_en_ las_terapias_cognitivo-conductuales
- Carrobes, J. A., & Buela-Casal, G. (1997). Técnicas psicofisiológicas. En G. Buela-Casal & J. C. Sierra (Eds.), *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones* (pp. 385-420). Madrid, España: Siglo XXI.
- Castro-Solano, A. (2003). ¿Son eficaces las psicoterapias psicológicas? *Psicodebate*, 3, 59-90. doi: [10.18682/pd.v3i0.503](https://doi.org/10.18682/pd.v3i0.503)
- Cautela, J. R., & Upper, D. (1976). The behavioral inven-

- tory battery: The use of self-report measures in behavioral analysis and therapy. En M. Harsen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (pp. 76-109). Oxford, Reino Unido: Pergamon.
- Clark, D. A., & Beck, A. T. (2012). *Terapia cognitiva para trastornos de ansiedad: Ciencia y práctica*. Bilbao: Desclée de Brouwer.
- De Ansorena-Cao, A., Cobo-Reinoso, J., & Romero-Cagigal, I. (1983). El constructo ansiedad en psicología: Una revisión. *Estudios de Psicología*, 4(16), 31-45. doi: [10.1080/02109395.1983.10821366](https://doi.org/10.1080/02109395.1983.10821366)
- Del Barrio, V. (1998). Evaluación de las características psicopatológicas. En R. Fernández-Ballesteros (Ed.), *Introducción a la Evaluación Psicológica* (Vol. II, pp. 120-159). Madrid, España: Pirámide.
- Derogatis, L. R. (1977). *Symptom Checklist -90-Revised*. PscyTESTS Dataset. doi: [10.1037/t01210-000](https://doi.org/10.1037/t01210-000)
- Echeburúa, E. (1996). Evaluación psicológica de los trastornos de ansiedad. En G. B. Buela-Casal, V. E. Caballo & J. C. Sierra (Eds.), *Manual de evaluación en psicología clínica y de la salud* (131-160). Madrid: Siglo XXI.
- Endler, N. S., Hunt, J. McV., & Rosenstein, A. J. (1962). An SR inventory of anxiousness. *Psychological Monographs: General and Applied*, 76(17), 1-33. doi: [10.1037/h0093817](https://doi.org/10.1037/h0093817)
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336-353. doi: [10.1037/1528-3542.7.2.336](https://doi.org/10.1037/1528-3542.7.2.336)
- Eysenck, M. W. (1988). Anxiety and attention. *Anxiety Research*, 1(1), 9-15. doi: [10.1080/10615808808248216](https://doi.org/10.1080/10615808808248216)
- Fernández-Ballesteros, R. (1994). Características básicas de la evaluación conductual. En R. Fernández-Ballesteros (Ed.), *Evaluación conductual hoy, un enfoque para el cambio en psicología clínica y de la salud* (pp. 85-110). Madrid, España: Pirámide.
- Fernández-Ballesteros, R. (2013a). Conceptos y modelos básicos. En R. Fernández-Ballesteros (Ed.), *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp. 27-60). Madrid, España: Pirámide.
- Fernández-Ballesteros, R. (2013b). La observación. En R. Fernández-Ballesteros (Ed.), *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp. 193-233). Madrid, España: Pirámide.
- Fernández-Ballesteros, R. (2013c). Los autoinformes. En R. Fernández-Ballesteros (Ed.), *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp. 235-274). Madrid, España: Pirámide.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2004). *Structured clinical interview for DSM-IV axis I disorders SCID-I: Clinician versión*. Washington: American Psychiatric.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2015). *Structured clinical interview for DSM-5, SCID-5: Clinician version*. Washington: American Psychiatric.
- Gómez-Penedo, J. M., & Roussos, A. (2012). ¿Cómo sabemos si nuestros pacientes mejoran? Criterios para la significancia clínica en psicoterapia: Un debate que se renueva. *Revista Argentina de Clínica Psicológica*, 21(2), 173-190. Recuperado de <http://www.clinicapsicologica.org.ar/index.php>
- González-Martínez, M. T. (1993). Aproximación al concepto de ansiedad en psicología: Su carácter complejo y multidimensional. *Aula*, 5, 9-22. Recuperado de <http://revistas.usal.es/index.php/0214-3402/index>
- Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32(1), 50-55. doi: [10.1111/j.2044-8341.1959.tb00467.x](https://doi.org/10.1111/j.2044-8341.1959.tb00467.x)
- Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1986). Evaluating the quality of behavioral assessment. En R. O. Nelson & S. C. Hayes, (Eds.), *Conceptual foundations of behavioral assessment* (pp. 463-503). New York, NY: Guilford Press.
- Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognitive Therapy and Research*, 4(4), 383-395. doi: [10.1007/bf01178214](https://doi.org/10.1007/bf01178214)
- James, W. (1884). What is an emotion? *Mind*, 9(34), 188-205. doi: [10.1093/mind/os-ix.34.188](https://doi.org/10.1093/mind/os-ix.34.188)

- James, W. (1890). *The Principles of Psychology* (Vol. I). New York, NY: Henry Holt. doi: [10.1037/10538-000](https://doi.org/10.1037/10538-000)
- Jensen, J. B. (1996). Los cuestionarios de autoinforme en la evaluación conductual. En G. B. Buena Casal, V. E. Caballo, & J. C. Sierra (Eds.), *Manual de evaluación en psicología clínica y de la salud* (pp. 109-127). Madrid: Siglo XXI.
- Keegan, E., & Holas, P. (2010). Cognitive-behavior therapy. Theory and practice. En Carlstedt, R. (Ed.), *Handbook of integrative clinical psychology, psychiatry and behavioral medicine. Perspectives, practices and research* (pp. 605-630). New York, NY: Springer.
- Landis, C., & Hunt, W. (1939). *The startle pattern*. Oxford, Reino Unido: Farrar & Rinehart.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. En P. J. Lang, R. F. Simons & M. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97-135). New York, NY: Routledge.
- Lang, P. J., Melamed, B. G., & Hart, J. (1970). A psychophysiological analysis of fear modification using an automated desensitization procedure. *Journal of Abnormal Psychology*, 76(2), 220-234. doi: [10.1037/h0029875](https://doi.org/10.1037/h0029875)
- Lange, C. G. (1885). *The Emotions*. Baltimore, MD: Williams and Wilkins.
- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., & Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, 12(5), 224-231. doi: [10.1016/s0924-9338\(97\)83296-8](https://doi.org/10.1016/s0924-9338(97)83296-8)
- Mandler, G., Mandler, J. M., & Ulville, E. T. (1958). Autonomic feedback: The perception of autonomic activity. *Journal of Abnormal & Social Psychology*, 56(3), 367-373. doi: [10.1037/h0048083](https://doi.org/10.1037/h0048083)
- Márquez, M. O. (2013). La entrevista. En R. Fernández-Ballesteros (Ed.), *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp.167-192). Madrid, España: Pirámide.
- Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy*, 23(5), 563-569. doi: [10.1016/0005-7967\(85\)90104-4](https://doi.org/10.1016/0005-7967(85)90104-4)
- Menéndez-Carrillo, F. X., & Maciá-Antón, D. (1994). Evaluación de los problemas de ansiedad. En R. Fernández-Ballesteros (Ed.), *Evaluación conductual hoy, un enfoque para el cambio en psicología clínica y de la salud* (pp. 426-483). Madrid, España: Pirámide.
- Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28(6), 487-495. doi: [10.1016/0005-7967\(90\)90135-6](https://doi.org/10.1016/0005-7967(90)90135-6)
- Miguel-Tobal, J. J., & Cano-Vindel, A. C. (1986). *ISRA Inventario de situaciones y respuestas de ansiedad*. Madrid: Tea.
- Miguel-Tobal, J. J. (1990). La ansiedad. En J. L. Pinillos & J. Mayor (Eds.), *Tratado de Psicología General. Motivación y Emoción*. Madrid, España: Alhambra Longman.
- Mohlman, J., & Gorman, J. M. (2005). The role of executive functioning in CBT: A pilot study with anxious older adults. *Behaviour Research and Therapy*, 43(4), 447-465. doi: [10.1016/j.brat.2004.03.007](https://doi.org/10.1016/j.brat.2004.03.007)
- Öst, L. G., Jerremalm, A., & Johansson, J. (1981). Individual response patterns and the effects of different behavioral methods in the treatment of social phobia. *Behaviour Research and Therapy*, 19(1), 1-16. doi: [10.1016/0005-7967\(81\)90107-8](https://doi.org/10.1016/0005-7967(81)90107-8)
- Öst, L. G., Johansson, J., & Jerremalm, A. (1982). Individual response patterns and the effects of different behavioral methods in the treatment of claustrophobia. *Behaviour Research and Therapy*, 20(5), 445-460. doi: [10.1016/0005-7967\(82\)90066-3](https://doi.org/10.1016/0005-7967(82)90066-3)
- Pacheco-Unguetti, A. P., Lupiáñez, J., & Acosta, A. (2009). Atención y ansiedad: Relaciones de la alerta y el control cognitivo con la ansiedad rasgo. *Psicológica*, 30(1), 1-25. Recuperado de <https://www.uv.es/psicologica>
- Pérez-Pareja, F. J. (1997). Autoinformes. En G. Buena-Casal

- & J. C. Sierra (Eds.), *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones* (pp. 297-314). Madrid, España: Siglo XXI.
- Quera, V., & Behar, J. (1997). La observación. En G. Buela-Casal & J. C. Sierra (Eds.), *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones* (pp. 315-341). Madrid, España: Siglo XXI.
- Sierra, J. C., Buela-Casal, G., & Fernández, M. A. (2004). La entrevista clínica. En G. Buela-Casal & J. C. Sierra (Eds.), *Manual de evaluación y tratamientos psicológicos* (pp. 41-68). Madrid: Biblioteca Nueva.
- Sierra, J. C., Ortega, V., & Zubeidat, I. (2003). Ansiedad, angustia y estrés: Tres conceptos a diferenciar. *Revista Mal-estar e Subjetividade*, 3(1), 10-59. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_serial&pid=1518-6148
- Silva, F. (1998). La entrevista. En R. Fernández-Ballesteros (Ed.) *Introducción a la Evaluación Psicológica* (Vol I, pp. 252-278). Madrid, España: Pirámide.
- Snaith, R. P., Baugh, S. J., Clayden, A. D., Husain, A., & Sipple, M. A. (1982). The Clinical Anxiety Scale: An instrument derived from the Hamilton Anxiety Scale. *British Journal of Psychiatry*, 141(5), 518-523. doi: [10.1192/bjp.141.5.518](https://doi.org/10.1192/bjp.141.5.518)
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1983). *Manual for the State-Trait Anxiety Inventory. STAI (Form Y), Self evaluation questionnaire*. Palo Alto, CA: Consulting Psychologists.
- Tryon, W. W. (1997). Técnicas de registro de la actividad motora. En G. Buela-Casal & J. C. Sierra (Eds.), *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones* (pp. 369-384). Madrid, España: Siglo XXI.
- Tyrer, P., Owen, R. T., & Cicchetti, D. V. (1984). The brief scale for anxiety: A subdivision of the comprehensive psychopathological rating scale. *Journal of Neurology, Neurosurgery & Psychiatry*, 47(9), 970-975. doi: [10.1136/jnnp.47.9.970](https://doi.org/10.1136/jnnp.47.9.970)
- Vila, J. (1998). Evaluación psicofisiológica. En R. Fernández-Ballesteros (Ed.), *Introducción a la Evaluación Psicológica* (Vol. II, pp. 335-370). Madrid, España: Pirámide.
- Vila, J. (2004). Evaluación neurofisiológica. En V. del Barrio (Ed.), *Evaluación psicológica aplicada a diferentes contextos*. Madrid, España: UNED.
- Vizcarro, C. (1998). Clasificación y criterios diagnósticos de la conducta anormal. En R. Fernández Ballesteros (Ed.), *Introducción a la Evaluación Psicológica* (Vol II, pp. 98-119). Madrid, España: Pirámide.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96(3), 465-490. doi: [10.1037//0033-2909.96.3.465](https://doi.org/10.1037//0033-2909.96.3.465)
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070. doi: [10.1037//0022-3514.54.6.1063](https://doi.org/10.1037//0022-3514.54.6.1063)
- Wittchen, H. U., & Boyer, P. (1998). Screening for anxiety disorders: Sensitivity and specificity of the Anxiety Screening Questionnaire (ASQ-15). *The British Journal of Psychiatry*, 173(S34), 10-17. doi: [10.1192/s000712500029346x](https://doi.org/10.1192/s000712500029346x)
- Wolpe, J. (1977). Inadequate behavior analysis: The Achilles heel of outcome research in behavior therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 8(1), 1-3. doi: [10.1016/0005-7916\(77\)90095-7](https://doi.org/10.1016/0005-7916(77)90095-7)
- Young, J. E. (1999). *Cognitive therapy for personality disorders: A schema focused approach*. (3ª ed.). Sarasota, FL: Professional Resource.
- Zigmond, A. S., & Snaith R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67(6), 361-370. doi: [10.1111/j.1600-0447.1983.tb09716.x](https://doi.org/10.1111/j.1600-0447.1983.tb09716.x)
- Zung, W. W. K. (1971). A rating instrument for anxiety disorders. *Psychosomatics*, 12(6), 371-379. doi: [10.1016/s0033-3182\(71\)71479-0](https://doi.org/10.1016/s0033-3182(71)71479-0)

Validación de la Escala de Esperanza Disposicional para Adultos en Paraguay

Validation of the Dispositional Adult Hope Scale in Paraguay

M. Alexandra Vuyk *¹, Gerónimo Codas¹

1 - Departamento de Psicología, Facultad de Filosofía y Ciencias Humanas, Universidad Católica Nuestra Señora de la Asunción, Asunción, Paraguay.

Introducción
Método
Resultados
Discusión
Conclusión
Referencias

Recibido: 24/09/2018 **Revisado:** 14/01/2019 **Aceptado:** 10/02/2019

Resumen

Hace algunas décadas, la esperanza disposicional se propuso como un constructo bifactorial. Sin embargo, evidencia reciente sugiere que podría entenderse mejor con un modelo unifactorial. En el presente estudio, mediante análisis factoriales confirmatorios se probó el ajuste de estructuras unifactoriales, con dos factores correlacionados, con dos factores más un factor de segundo orden, y bifactor, de la Escala de Esperanza Disposicional para Adultos (Snyder et al., 1991), traducida al español y adaptada a la población paraguaya. Se realizó el cuestionario a 219 estudiantes universitarios de Paraguay (edad promedio de 23 años; DE = 3.717; 55 participantes de sexo masculino y 163 de sexo femenino). Los resultados indican un excelente ajuste para los tres modelos, considerándose preferible el modelo bifactor debido a que los índices complementarios indican que la escala es principalmente unidimensional. La escala demuestra propiedades psicométricas adecuadas para su uso en poblaciones latinoamericanas.

Palabras clave: *Análisis factorial confirmatorio, modelos bifactor, corrección Satorra-Bentler, Escala de Esperanza Disposicional para Adultos, adaptación transcultural*

Abstract

Trait Hope has been proposed decades ago as a bifactorial construct. Nevertheless, recent evidence has supported the idea of a unifactorial model. In this study, we present results for confirmatory factor analyses in which we assess fit for models that include unifactorial, two correlated factors, two factors with a higher order factor, and bifactor structures of the Adult Trait Hope Scale's (Snyder et al., 1991) adaptation to a Spanish-speaking Paraguayan population. The sample is constituted by 219 college students in Paraguay (mean age 23 years old; SD = 3.717; 55 of the students are male and 163 are female). Results show excellent fit for all three models, with a preference for the bifactor model as complementary indices show the scale is primarily unidimensional. The scale shows adequate psychometric properties for its use in Latin American populations.

Keywords: *Confirmatory factor analysis, bifactor models, Satorra-Bentler correction, Adult Trait Hope Scale, cross-cultural adaptation*

Introducción

Hace casi tres décadas, Snyder et al. (1991) plantearon el concepto de *esperanza* como un dispositivo cognitivo basado en dos conceptos recíprocamente determinados: *agencia* y *medios*. Ambos conceptos constituyen la base del modelo propuesto por dichos autores: la *agencia*, por un lado, puede entenderse como la determinación o motivación hacia una meta; los *medios*, por otro, son la serie de recursos o estrategias a través de las cuales el individuo pretende la obtención de dicha meta. Además, los autores no conceptualizan la esperanza como un fenómeno que se da distintamente dependiendo de cada meta y contexto en particular, sino como un rasgo o disposición que define la actitud con que los individuos asumen sus objetivos.

De acuerdo con la idea de Snyder (2002), la esperanza es una forma de pensar que se aprende en la infancia a través de la consecución y no consecución de metas. Los individuos con altos y bajos indicadores de esperanza asumen las tareas de modos cualitativamente distintos, y dan distintas respuestas emocionales ante los resultados. Las emociones que aparecen durante el proceso de fijar una determinada meta y las posibles estrategias para aproximarse a la misma retroalimentan al dispositivo de la esperanza, determinando las vivencias que se tendrá en experiencias futuras.

Relacionando la esperanza con otros constructos, Snyder, Feldman, Taylor, Schroeder y Adams (2000) propusieron que la esperanza amplificaría la satisfacción en los individuos. En esta línea, se ha presentado evidencia (Yarcheski & Mahon, 2014) de la relación positiva entre la esperanza y la satisfacción con la vida. Recientemente, el concepto de esperanza ganó relevancia en el estudio de la orientación vocacional, donde se toma como un factor fundamental en las conductas de exploración (Hirschi, Abessolo, & Froi-

devaux, 2015).

Estructura Factorial de la Escala de Esperanza Disposicional

La Escala de Esperanza Disposicional para Adultos (Snyder et al., 1991), diseñada para recabar datos sobre la teoría de Snyder, consiste en doce ítems de los cuales ocho están orientados a evaluar ambas dimensiones de la esperanza disposicional y cuatro ítems sirven como distractores. Además, se ha diseñado una Escala de Esperanza-Estado (Snyder et al., 1996) y una Escala de Esperanza para Niños (Snyder et al., 1997). Sin embargo, estas no serán tenidas en cuenta en el presente estudio.

Originalmente, la escala fue validada en estudiantes universitarios de la carrera de psicología y pacientes de dos instituciones en los Estados Unidos (Snyder et al., 1991). Para asegurar la validez de contenido, Snyder y su equipo primeramente escribieron 45 ítems que reflejaran el concepto hipotético de la esperanza. Estos 45 ítems fueron administrados a estudiantes universitarios para luego condensarlos en un instrumento más breve. Los ítems con bajas correlaciones ítem-total fueron descartados, y se mantuvieron 14 ítems con correlaciones ítem-total en niveles aceptables. Luego, Snyder personalmente seleccionó los cuatro ítems que mejor reflejaran el componente de *agencia* y los cuatro ítems que mejor reflejaran el componente de *medios*.

Aparte de presentar las evidencias respecto de este estudio, Snyder et al. (1991) recopilaron distintas evidencias de validez y confiabilidad para la misma. El coeficiente alfa de Cronbach para la escala total fue calculado en varias muestras; los resultados oscilaron entre .74 y .84. Para la escala de *agencia*, los coeficientes alfa fueron de entre .71 y .76, mientras que para *medios* fue-

ron de entre .63 y .80. Se calculó la confiabilidad test-retest con un intervalo de 3 y 8 semanas ($r = .85$; $p < .001$; Anderson, 1988, como se citó en Snyder et al., 1991; y $r = .73$; $p < .001$; Harney, 1989, como se citó en Snyder et al., 1991; respectivamente) y de 10 semanas con dos muestras distintas ($r = .76$; $p < .001$ y $r = .82$; $p < .001$; Gibb, 1990, como se citó en Snyder, 1991; y Yoshinobu, 1989, como se citó en Snyder, 1991, respectivamente).

Además, se halló evidencia de validez convergente con escalas de optimismo (correlaciones de .60 y .50, con valores $p < .005$ en dos estudios), expectativas generalizadas de éxito (correlaciones de .55 y .54, con valores $p < .005$ en dos estudios), deseabilidad de control ($r = .54$; $p < .005$), resolución de problemas ($r = -.62$; $p < .005$; puntajes más bajos de esta escala reflejan mayor resolución de problemas), autoestima ($r = .58$; $p < .005$) y dos escalas de deseabilidad social (correlaciones de .30 y .28; $p < .005$). Se halló también evidencia de validez divergente con escalas de desesperanza ($r = -.51$; $p < .005$) y depresión ($r = -.42$; $p < .005$; Gibb, 1990, como se citó en Snyder, 1991; Holleran & Snyder, 1990, como se citó en Snyder, 1991; Irving, Crenshaw, Snyder, Francis, & Gentry, 1990, como se citó en Snyder, 1991). Más evidencia de validez concurrente se halló con las subescalas del MMPI y de las Frases Incompletas de Rotter, obteniendo los siguientes resultados: hipocondriasis ($r = -.30$; $p < .001$), depresión ($r = -.60$; $p < .001$), histeria ($r = -.35$; $p < .001$), desviación psicopática ($r = -.43$; $p < .001$), masculinidad-femineidad ($r = -.13$; ns), paranoia ($r = -.34$; $p < .001$), psicastenia ($r = -.52$; $p < .001$), esquizofrenia ($r = -.46$; $p < .001$), hipomanía ($r = -.08$; ns), introversión social ($r = -.59$; $p < .001$), completamiento de frases ($r = .63$; $p < .001$). Evidencia de validez discriminante se presentó con respecto a dos subescalas de una escala

de autoconciencia ($r = .06$ y $-.03$, respectivamente, ns; Gibb, 1990, como se citó en Snyder, 1991).

En cuanto a la validez de constructo, varios estudios confirmatorios experimentales han presentado buena evidencia con respecto al comportamiento de la variable frente a la presencia de estresores, cantidad de objetivos, la dificultad preferida de los objetivos, la evaluación de objetivos vitales, los objetivos académicos, evaluación de consecución de objetivos y la consecución de objetivos (Yoshinobu, 1989; Langelle, 1989; Harris, 1988; Sigmon & Snyder, 1990; Anderson, 1988; Harney, 1989, todos citados en Snyder et al., 1991). Estos estudios responden a lo hipotetizado sobre la influencia de la esperanza.

Con respecto a la estructura factorial, análisis factoriales favorecieron una estructura de dos factores correlacionados en el cuestionario, en concordancia con los conceptos de agencia y medios (Snyder et al., 1991). Por otro lado, estudios más recientes (Brower, Meijer, Weekers, & Baneke, 2008) proporcionaron evidencia de que una estructura unidimensional presenta un mejor ajuste para la escala, además de cuestionar la validez incremental que provee la subescala de medios. Por el contrario, Gana, Daigre y Ledrich (2012) apoyaron la estructura de dos factores, y en su estudio, las escalas del cuestionario presentaban una correlación fuerte entre ellas.

Galiana, Oliver, Sancho y Tomás (2014) desarrollaron lo que aparentemente es la primera traducción de la escala al español. Dicho estudio presentó evidencia de análisis factoriales confirmatorios para modelos unifactorial, multidimensional no jerárquico, de segundo orden y de dos factores, hallando mejor ajuste en el caso de la estructura unidimensional. Además, Galiana et al. (2014) señalan que la consistencia interna del cuestionario fue mayor cuando fue tomado como unidimensional. Estos resultados alimentan el de-

bate sobre la composición de la esperanza.

Al haber resultados contradictorios acerca de la unidimensionalidad vs. multidimensionalidad, es plausible que la Escala de Esperanza se ajuste a un modelo bifactor. Los modelos bifactor evalúan si un conjunto de ítems representa una dimensión general subyacente además de factores específicos (Flores-Kanter, Domínguez-Lara, Trógolo, & Medrano, 2018). De ser este el caso, la esperanza se conceptualizaría como un constructo unidimensional que incluye factores específicos como agencia y medios.

En Latinoamérica se ha utilizado el instrumento de Snyder para investigación en el contexto de la psicología del deporte (Guillén & Angulo, 2016), psicología positiva (Bustos, Oliver, & Galiana, 2015) y la psicología organizacional (Zárate-Torres & Acosta-Prado, 2015). Para el presente estudio, se elaboró una nueva traducción del instrumento, adaptada a la población paraguaya (ver Tabla 1). Se buscaron traducciones funcionales y no literales, con el fin de que resultaran culturalmente congruentes según las recomendaciones de adaptaciones de cuestionarios de Fernández, Pérez, Alderete, Richaud y Fernández-Liporace (2010).

Buscando aportar nuevos datos a la discusión, nuestra pregunta de investigación indaga si

la Escala de Esperanza Disposicional para Adultos en la muestra paraguaya se ajusta a la estructura de dos factores correlacionados de agencia y medios, de dos factores de agencia y medios con esperanza como factor de segundo orden, unifactorial de esperanza, o bifactor con esperanza como factor general y agencia y medios como factores específicos.

Método

Participantes

Se recogieron los datos en universidades públicas y privadas de Paraguay (ver Tabla 1). Un total de 219 estudiantes de grado, mayores de 18 años (55 de sexo masculino, 163 de sexo femenino, uno sin respuesta) y de distintas carreras, completaron los cuestionarios. La edad promedio fue de 23 años de edad ($DE = 3.743$). Los estudiantes cursaban materias de entre el primer y el séptimo año de sus respectivas carreras. Los participantes fueron informados de que eran libres de completar los cuestionarios, y que no se aplicaría ninguna penalización en caso de no participar. No se ofrecieron incentivos por completar los instrumentos.

La aplicación del instrumento a 164 estu-

Tabla 1

Datos sociodemográficos de la muestra ($N = 219$).

Característica	n	%
Sexo*		
Femenino	163	74
Masculino	55	25
Estado Civil*		
Soltero/a	190	87
Viviendo en pareja	6	3
Casado/a	22	10
Número de hijos*		
Ninguno	190	87
Uno	11	5
Dos	5	2

Nota. * Estas variables presentan datos perdidos en el sistema.

diantes de la muestra se realizó en las aulas durante el horario de clases. Se solicitó autorización a las instituciones y a cada docente para hacer uso de los últimos 30 minutos de clase para este fin. Al terminar, los participantes depositaron sus cuestionarios completados en una mesa designada antes de retirarse del aula. Por otro lado, 55 estudiantes completaron una versión de los cuestionarios diseñada en *Qualtrics*. Se alcanzó a dichos estudiantes por medio de un muestreo tipo bola de nieve compartiendo los cuestionarios en redes sociales como *Facebook* y *Twitter*. Antes del levantamiento de los datos, todos los participantes leyeron y firmaron formularios de consentimiento informado. El procedimiento de este estudio fue aprobado previamente por el comité de ética de la Universidad de Kansas, afiliación institucional de la primera autora en aquel momento.

Análisis de poder y tamaño muestral. Varios análisis de poder fueron realizados utilizando la calculadora de [Preacher y Coffman \(2006\)](#) para asegurar un tamaño muestral adecuado para este estudio. Los cálculos de grados de libertad tuvieron en cuenta los modelos hipotéticos a contrastar. Para comprobar el poder para evaluar el ajuste total del modelo, se realizó el análisis de ajuste pobre de [MacCallum, Browne y Sugawara \(1996\)](#). El análisis de ajuste pobre nos permite saber si, dado el tamaño muestral, se podría rechazar una hipótesis de ajuste pobre en la muestra si el ajuste de la población fuera bueno y viceversa. Basado en los modelos, los grados de libertad van entre 12 y 20. Se especificó una hipótesis nula de pobre ajuste con un valor de RMSEA de .10 o mayor, y una hipótesis alternativa de buen ajuste con un valor de RMSEA de .02 o menor. Para 12 grados de libertad, .80 de poder, y .05 de valor alfa, el análisis de poder indicó que el tamaño muestral mínimo sería de $N = 166$; para 20 grados de libertad, .80 de poder, y .05 de alfa, el tamaño muestral

mínimo sería de $N = 119$. Por lo tanto, el tamaño muestral obtenido de 219 tiene suficiente poder para los análisis realizados.

Instrumentos

Cuestionario sociodemográfico. Incluyó preguntas relacionadas con la edad, sexo, nombre de la universidad, carrera, años en la carrera, si era el primer miembro de su familia en asistir a la universidad, grado de seguridad en la elección de la carrera, ocupación de los padres, nivel de instrucción, estado civil, número de hijos, y quién había elegido la carrera que se encontraban estudiando (ej., si la habían elegido ellos o sus padres). Este cuestionario fue generado por la primera autora en colaboración con otros investigadores para estudios en la misma línea.

Escala de Esperanza Disposicional para Adultos (Snyder et al., 1991). Consta de 12 reactivos dispuestos en dos subescalas: cuatro para la subescala de Agencia, cuatro para la de Medios y cuatro ítems adicionales que actúan como distractores. Las opciones de respuesta se presentan en una escala Likert de ocho puntos, que va desde 1 (*Definitivamente falso*) hasta 8 (*Definitivamente verdadero*). Los puntajes en las subescalas son aditivos y forman un puntaje total. El coeficiente alfa de Cronbach para la presente muestra es de .78 para la subescala de Agencia, de .69 para la subescala de Medios y de .83 para la escala total. Como el instrumento original de [Snyder et al. \(1991\)](#) fue redactado en inglés, para el presente estudio utilizamos el procedimiento de traducción y re-traducción para adaptarlo a la población hispanoparlante en Paraguay, siguiendo los estándares de la comunidad científica enfatizados en [Fernández et al. \(2010\)](#). Un experto en español e inglés con conocimientos de psicología tradujo la

versión original de los instrumentos al español. Posteriormente, otro experto en ambos idiomas tradujo la versión en español de nuevo al inglés. Un tercer experto en ambos idiomas comparó ambas versiones para asegurarse de que los reactivos fueran análogos y mantuvieran su significado.

Plan de Análisis de Datos

Primero, se realizaron análisis descriptivos de los ítems y las subescalas que componen la Escala de Esperanza. Se midió asimetría y curtosis tomando como indicadores de normalidad su transformación a puntajes z dividiendo los puntajes por su error estándar y verificando su significancia.

Se realizaron análisis factoriales confir-

Tabla 2
Reactivos de la Escala de Esperanza Traducidos al Español.

Número de reactivo	Enunciado del reactivo
1	Puedo pensar en muchas maneras de salir de un embrollo.
2	Persigo mis metas con energía.
3	Me siento cansado/a la mayoría del tiempo.
4	Hay muchas maneras de solucionar cualquier problema.
5	Me siento rápidamente vencido/a en una discusión.
6	Puedo pensar en muchas maneras de conseguir las cosas que son importantes para mi vida.
7	Me preocupo por mi salud.
8	Incluso cuando otros se sienten desanimados, yo sé que puedo encontrar una manera de solucionar el problema.
9	Mis experiencias pasadas me han preparado bien para mi futuro.
10	He sido bastante exitoso/a en la vida.
11	Usualmente me encuentro preocupado/a por algo.
12	Cumplo las metas que me trazo.

matorios en el paquete lavaan del programa *R* (Rosseel, 2012) mediante estimación de máxima probabilidad; además, se utilizó la estimación robusta de máxima probabilidad con la corrección de chi-cuadrado Satorra-Bentler, al haber desviación de la normalidad en la distribución de los puntajes según recomendación de Finney y DiStefano (2013), quienes además sugieren la presentación de los resultados con ambos métodos. Estos análisis compararon cuatro modelos: la esperanza como un constructo de dos factores correlacionados, la esperanza como un constructo de dos factores con un factor de segundo orden, la esperanza como un constructo unifactorial y la esperanza como un constructo bifactor.

Los modelos en el análisis factorial confirmatorio también se llaman modelos de medición,

y definen a los constructos latentes mediante múltiples indicadores corrigiendo así el error de medición (Little, 2013). De este modo, pueden establecer la evidencia de validez basada en el contenido de estos indicadores separando a los confiables de los no confiables. Los parámetros poblacionales en modelos de ecuaciones estructurales como los análisis factoriales confirmatorios no incluyen sesgo y por ende son más exactos y más generalizables (Little, 2013).

La varianza factorial se fijó mediante el método de configuración de la escala (*scale setting*; Little, 2013). Las estadísticas de bondad de ajuste siguieron las sugerencias de Hu y Bentler (1999) y Little (2013) acerca de un ajuste aceptable con $CFI > .90$, $RMSEA < .08$ y $SRMR < .11$, o ajuste excelente con $CFI > .95$, $RMSEA < .05$ y $SRMR$

< .06. Para comparar los modelos, se tomaron en cuenta las convenciones para los índices de ajuste; en caso de ser apropiadas, se utilizó la regla de [Cheung y Rensvold \(2002\)](#) de continuar evaluando el siguiente modelo si el cambio en CFI fuera menor a < .01.

Los datos perdidos en las variables de interés estuvieron entre 0 y 1.4% (.004% de los puntos totales) y fueron considerados falta completamente al azar o MCAR (*missing completely at random*; [Little, 2013](#)). Fueron estimados en los análisis de máxima probabilidad mediante el uso del método de información completa (*full estimation maximum likelihood* o FIML en lavaan). Los análisis con la corrección robusta de Satorra-Bentler no admiten el uso de la información completa, y en ese caso la muestra final para estos análisis fue de 212 participantes debido a la eliminación de los casos con datos perdidos.

Además, se evaluaron índices complementarios para estimar la adecuación del modelo bifactor, como la varianza total explicada (ECV), la varianza total explicada por ítem (IECV), el índice de replicabilidad del constructo (H-index),

el porcentaje de correlaciones no contaminadas (PUC) y los coeficientes de confiabilidad Omega. Para el cálculo de dichos índices, se utilizaron las calculadoras disponibles en línea de [Dueber \(2017\)](#) y [Hammer \(2016\)](#). Estos índices son especialmente importantes para una interpretación apropiada, ya que los modelos bifactor tienden a presentar mejor ajuste que los análisis factoriales confirmatorios solamente debido a la manera en la que se los especifica ([Flores-Kanter et al., 2018](#)).

Resultados

Los puntajes medios en cada ítem (ver Tabla 3) oscilan entre los tres y los siete puntos. Los desvíos estándar se encuentran entre uno y dos puntos para cada caso.

Los valores de asimetría y curtosis observados en la Tabla 3, transformados a puntajes z , oscilan entre -5.95 y -4.61 para asimetría y 1.15 y 4.17 para curtosis, dándonos asimetría y curtosis significativas ($p < .001$). Por lo tanto, se presentan tanto los análisis con estimación de máxima

Tabla 3
Indicadores Descriptivos.

Ítems	M	DE	Asimetría	Curtosis	α de Cronbach
1 - Medios	6.48	1.22	-.89	.86	
2 - Agencia	6.49	1.23	-.79	.53	
3 - Distractor	5.00	1.87	-.36	-.88	
4 - Medios	6.81	1.28	-1.50	2.89	
5 - Distractor	3.38	2.04	.58	-.79	
6 - Medios	6.70	1.12	-.79	.09	
7 - Distractor	6.35	1.60	-1.11	.91	
8 - Medios	6.43	1.33	-1.24	2.22	
9 - Agencia	6.60	1.41	-1.31	2.09	
10 - Agencia	5.96	1.41	-.89	.71	
11 - Distractor	5.86	1.56	-.80	.41	
12 - Agencia	6.30	1.26	-.74	.39	
Escala de Agencia	25.35	4.13	-.99	1.38	.69
Escala de Medios	26.48	3.54	-.76	.38	.78
Total	51.89	6.91	-.90	1.09	.83

Tabla 4
Matriz de Covarianzas y Correlaciones entre Reactivos.

Reactivos	Medios				Agencia			
	1	4	6	8	2	9	10	12
1	1.445	.432	.565	.688	.521	.560	.458	.454
4	<i>.281</i>	1.642	.311	.582	.453	.536	.688	.465
6	<i>.419</i>	<i>.216</i>	1.256	.641	.559	.360	.477	.437
8	<i>.438</i>	<i>.347</i>	<i>.437</i>	1.709	.749	.808	.736	.702
2	<i>.351</i>	<i>.285</i>	<i>.403</i>	<i>.463</i>	1.531	.751	.799	.848
9	<i>.327</i>	<i>.294</i>	<i>.225</i>	<i>.434</i>	<i>.426</i>	2.033	.953	.842
10	<i>.268</i>	<i>.378</i>	<i>.300</i>	<i>.397</i>	<i>.455</i>	<i>.471</i>	2.014	.884
12	<i>.302</i>	<i>.290</i>	<i>.311</i>	<i>.429</i>	<i>.548</i>	<i>.472</i>	<i>.498</i>	1.567

Nota. Las covarianzas se presentan en la diagonal y por encima de la diagonal. Las correlaciones se presentan debajo de la diagonal, en cursiva y negrita.

probabilidad, así como los análisis factoriales confirmatorios con chi-cuadrado de Satorra-Bentler que tienen en cuenta la desviación de la normalidad. En la Tabla 4 puede apreciarse la matriz de correlaciones y covarianzas, encontrándose correlaciones entre los reactivos que componen al instrumento.

Los datos (Tabla 5) indican la bondad de ajuste de los modelos con análisis de máxima probabilidad y chi-cuadrado de Satorra-Bentler, que presentan resultados similares. Según estos análisis, tanto la estructura de dos factores correlacionados, la estructura de dos factores con un factor de segundo orden, la unifactorial y la bifactor se ajustan de manera excelente, siguiendo los criterios de interpretación de ajuste propuestos por Hu y Bentler (1999) y Little (2013).

El CFI del modelo de dos factores correlacionados, del modelo de dos factores con un factor de segundo orden y del modelo bifactor fueron idénticos en los análisis de Satorra-Bentler, y prácticamente idénticos en los análisis con estimación de máxima probabilidad. No se pudo calcular índices de ajuste para el modelo bifactor con estimación de máxima probabilidad por fallas de convergencia luego de 10000 iteraciones. En

comparación con el modelo unifactorial, el cambio del CFI fue mayor a .01 para ambas estimaciones, lo que indica que los modelos unifactoriales tienen un peor ajuste (Cheung & Rensvold, 2002).

Al presentar excelente ajuste, los modelos de dos factores correlacionados, dos factores con un factor de segundo orden y bifactor, fue necesario recurrir a los índices complementarios para poder elucidar la unidimensionalidad o bidimensionalidad de la Escala de Esperanza Disposicional.

La varianza común explicada total del modelo bifactor se computó con la calculadora en línea de Dueber (2017) y fue de ECV = .768 para el factor general, ECV-S = .273 para el factor específico de *agencia* y ECV-S = .177 para el factor específico de *medios*. El porcentaje de correlaciones no contaminadas o PUC fue de .571, que indica las covarianzas o correlaciones que solamente reflejan variación de la dimensión general y no están contaminadas con multidimensionalidad (Flores-Kanter et al., 2018).

Además, se calculó la varianza común explicada del ítem o IECV, que indica si la varianza en las respuestas a un ítem se puede atribuir a la

Tabla 5
Índices de Bondad de Ajuste de los Análisis Factoriales Confirmatorios.

Modelo	<i>df</i>	χ^2	<i>p</i> χ^2	CFI	AIC	RMSEA	SRMR	Modelo de Comparación	Δ CFI
1.1 Dos factores correlacionados, estimación de máxima probabilidad (<i>N</i> = 219)	19	23.853	.202	.989	5393.119	.034	.033		
1.2 Dos factores correlacionados, Satorra-Bentler (<i>N</i> = 212)	19	15.475	.692	1.000	5219.461	.000	.034		
2.1 Dos factores con un factor de segundo orden, estimación de máxima probabilidad (<i>N</i> = 219)	18	23.853	.160	.987	5395.119	.039	.033	2.1 vs 1.1	.002
2.2 Dos factores con un factor de segundo orden, Satorra-Bentler (<i>N</i> = 212)	18	15.599	.621	1.000	5221.461	.000	.033	2.2 vs 1.2	.000
3.1 Unifactorial, estimación de máxima probabilidad (<i>N</i> = 219)	20	40.731	.004	.954	5407.997	.069	.044	3.1 vs 1.1	.035
3.2 Unifactorial, Satorra-Bentler (<i>N</i> = 212)	20	28.304	.102	.968	5235.857	.044	.044	3.2 vs 1.2	.032
4.1 Bifactor, estimación de máxima probabilidad (<i>N</i> = 219)*	-	-	-	-	-	-	-	-	-
4.2 Bifactor, Satorra-Bentler (<i>N</i> = 212)	12	10.752	.550	1.000	5209.093	.000	.027	4.2 vs 1.2	.000

Nota. No fue posible calcular índices de ajuste para el modelo 4.1, bifactor con estimación de máxima probabilidad, debido a que dicho modelo no logró convergencia luego de 10000 iteraciones y por ende no es confiable.

varianza en la dimensión general; puede ser útil para evaluar la unidimensionalidad e inclusive extraer un conjunto de ítems que reflejen el contenido de la dimensión general (Stucky & Edelen, 2015). La IECV en la Escala de Esperanza estuvo entre .534 y .965, con cinco ítems por encima del valor de .80 que sugiere que dicho ítem refleja la dimensión general (ver Tabla 6).

Se calculó el índice de replicabilidad del constructo o H-index (Rodríguez, Reise, & Haviland, 2016) utilizando la calculadora de Hammer (2016), y se obtuvo un índice de .942 para

el factor general de esperanza, .578 para el factor específico de *agencia* y .371 para el factor específico de *medios*; para el modelo de dos factores correlacionados, el H-index fue de .96 tanto para la subescala de Agencia como para la de Medios. Valores por encima de .80 indican que la variable latente está bien definida y el constructo podrá ser replicado en estudios posteriores (Rodríguez et al., 2016).

La confiabilidad fue estudiada con los coeficientes Omega (Flores-Kanter et al., 2018).

Tabla 6
Carga Factorial e IECV de la Escala de Esperanza en el Modelo Bifactor.

Ítems	Factor General	Factor Específico Medios	Factor Específico Agencia	IECV
1 - Medios	.657	.326		.802
4 - Medios	.662	-.124		.966
6 - Medios	.583	.545		.534
8 - Medios	.938	.179		.965
2 - Agencia	.782		.386	.804
9 - Agencia	.817		.393	.812
10 - Agencia	.830		.445	.777
12 - Agencia	.726		.659	.548

El coeficiente Omega (ω) es un estimativo de la confiabilidad interna del constructo multidimensional, basado en el modelo especificado, en donde para el factor general se consideran todos los ítems, y en este estudio fue de .952; para el Omega de subescalas (ω_s) se consideran solamente los ítems que cargan en ese factor específico, hallándose $\omega_s = .960$ para la subescala de Agencia y $\omega_s = .859$ para la subescala de Medios. El coeficiente Omega jerárquico u OmegaH (ω_h) indica la proporción de la varianza que puede ser atribuida a diferencias individuales en el factor general; en este estudio, $\omega_h = .848$. El coeficiente Omega jerárquico de subescalas u OmegaHS (ω_{hs}) indica la proporción de la varianza correspondiente a una subescala luego de controlar la varianza atribuida al factor general, y se halló $\omega_{hs} = .252$ para la subescala de Agencia y $\omega_{hs} = .083$ para la subescala de Medios.

Discusión

El presente estudio se centró en la validación factorial de la Escala de Esperanza Disposicional en Adultos (Snyder et al., 1991). Se han propuesto distintos modelos de ajuste para dicho instrumento, desde el unifactorial (Brower et al., 2008; Galiana et al., 2014) hasta el de dos factores (Snyder et al., 1991; Gana et al., 2012). En este trabajo, se aplicó un análisis factorial confir-

matorio para probar el ajuste con un modelo unifactorial, un modelo de dos factores correlacionados, un modelo de dos factores con un factor de segundo orden y un modelo bifactor.

Los datos muestran que el instrumento presenta ajuste óptimo con los tres modelos. Sin embargo, a nuestro criterio, el modelo con mejor ajuste es el bifactor. Las subescalas de Agencia y Medios presentan puntajes más bajos en el alfa de Cronbach (.78 en la subescala de Agencia, .69 en la subescala de Medios) en comparación con la escala total con un alfa de Cronbach de .83. El ajuste unifactorial también técnicamente presenta un ajuste excelente en esta muestra, aunque ligeramente inferior a los modelos de dos factores y bifactor.

Como sugieren Reise, Scheines, Widaman, y Haviland (2013), cuando los valores de PUC son inferiores a .80 (en este estudio, PUC = .571), la ECV general mayor a .60 (en este estudio, ECV = .768) y OmegaH del factor general mayor a .70 (en este estudio, $\omega_h = .848$) cierta multidimensionalidad existe, pero no es lo suficientemente severa para descalificar la interpretación de la escala como principalmente unidimensional. Por ende, según estos resultados, la esperanza es un constructo preponderantemente unidimensional con ciertas características específicas.

Estos resultados apoyan la estructura teórica que se ha propuesto para la esperanza (Snyder et

al., 1991; Gana et al., 2012) como un constructo general incluyendo dos factores específicos, y explican por qué otros estudios pudieron hallar apoyo para una estructura unidimensional (Galiana et al., 2014). Al ajustarse la Escala de Esperanza Disposicional a un modelo bifactor, se resalta la unidimensionalidad primaria del constructo de esperanza, a pesar de la presencia de cierta multidimensionalidad; esto implica que estas escalas no son independientes, sino que ambas son parte específica de la esperanza general. A nivel teórico, esto podría implicar constante interacción y retroalimentación entre ambos componentes específicos de la esperanza como estado general.

Inclusive, basándonos en los IECV, se podría crear una versión breve de la Escala de Esperanza Disposicional que refleje el constructo de la esperanza como unidimensional (Stucky & Edehlen, 2015); esta versión se presenta en la Tabla 7.

Los resultados también indican que la traducción de la escala presenta validez factorial y consistencia interna, lo que favorece su utilización en investigación y práctica con muestras paraguayas y potencialmente de toda Latinoamérica. A partir de estas conclusiones, la medición de la esperanza con el presente instrumento podrá ser incluida en modelos como factor o como mediador de otros factores. En resumen, los resultados indican que la escala es sumamente útil para la medición de la esperanza disposicional.

Como limitaciones del estudio, se destaca que los participantes fueron encontrados en ámbitos universitarios, lo que quizás limitaría las conclusiones a estratos sociales con un nivel académico avanzado. Al ser un muestreo por conveniencia y bola de nieve, podría no ser generalizable a la población universitaria general. Además, gran parte de la población paraguaya habla solamente idioma guaraní, teniendo conocimientos no siempre óptimos del español. Por ese motivo, la escala no podría ser aplicada en esos estratos.

Conclusión

El análisis factorial confirmatorio realizado en este estudio permite concluir que la Escala de Esperanza Disposicional en Adultos (Snyder et al., 1991) se ajusta de mejor manera a un modelo bifactor; por lo tanto, la esperanza es un factor general que engloba agencia y medios como factores específicos. La traducción de esta escala al español presenta validez factorial y consistencia interna para muestras paraguayas, lo que permite su utilización específicamente en individuos con buen manejo del idioma español. Replicaciones de la estructura factorial en otras poblaciones latinoamericanas servirán para generalizar la utilidad del instrumento. Al tener un instrumento que presenta evidencias de validez, se puede conti-

Tabla 7
Reactivos de la Escala de Esperanza para Versión Breve.

Número original de reactivo	Enunciado del reactivo
1	Puedo pensar en muchas maneras de salir de un embrollo.
2	Persigo mis metas con energía.
4	Hay muchas maneras de solucionar cualquier problema.
8	Incluso cuando otros se sienten desanimados, yo sé que puedo encontrar una manera de solucionar el problema.
9	Mis experiencias pasadas me han preparado bien para mi futuro.

nuar evaluando la relación entre características de personalidad o rasgos conductuales que pudieran estar relacionados con la esperanza, o moderados o mediados por esta.

Referencias

- Brower, D., Meijer, R. R., Weekers, A. M., & Baneke, J. J. (2008). On the dimensionality of the Dispositional Hope Scale. *Psychological Assessment, 20*(3), 310-315. doi: [10.1037/1040-3590.20.3.310](https://doi.org/10.1037/1040-3590.20.3.310)
- Bustos, V., Oliver, A., & Galiana, L. (2015). Validación del autoconcepto forma 5 en universitarios peruanos: Una herramienta para la psicología positiva. *Psicología: Reflexão e Crítica, 28*(4), 690-697. doi: [10.1590/1678-7153.201528406](https://doi.org/10.1590/1678-7153.201528406)
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi: [10.1207/s15328007sem0902_5](https://doi.org/10.1207/s15328007sem0902_5)
- Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models [software de cálculo]. Disponible en <https://doi.org/10.13023/edp.tool.01> , https://uknowledge.uky.edu/edp_tools/1/
- 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. *Revista Evaluar, 10*(1), 60-74. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Flores-Kanter, P. E., Dominguez-Lara, S., Trógolo, M. A., & Medrano, L. A. (2018). Best practices in the use of bifactor models: Conceptual grounds, fit indices and complementary indicators. *Revista Evaluar, 18*(3), 44-48. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. En G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2ª ed., pp. 439-492). Charlotte, NC: Information Age.
- Galiana, L., Oliver, A., Sancho, P., & Tomás, J. M. (2014). Dimensionality and validation of the Dispositional Hope Scale in a Spanish sample. *Social Indicators Research, 120*(1), 297-308. doi: [10.1007/s11205-014-0582-1](https://doi.org/10.1007/s11205-014-0582-1)
- Gana, K., Daigre, S., & Ledrich, J. (2012). Psychometric properties of the French version of the Adult Dispositional Hope Scale. *Assessment, 20*(1), 114-118. doi: [10.1177/1073191112468315](https://doi.org/10.1177/1073191112468315)
- Guillén, F., & Angulo, J. (2016). Análisis de rasgos de personalidad positiva y bienestar psicológico en personas mayores practicantes de ejercicio físico vs no practicantes. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte, 11*(1), 113-122. Recuperado de <http://www.riped-online.com/index.php/riped/inicio>
- Hammer, J. H. (2016). Construct Replicability Calculator: A Microsoft Excel-based tool to calculate the Hancock and Mueller (2001) H index [software de cálculo]. Recuperado de <http://drjosephhammer.com> el 06/03/2019
- Hirschi, A., Abessolo, M., & Froidevaux, A. (2015). Hope as a resource for career exploration: Examining incremental and cross-lagged effects. *Journal of Vocational Behavior, 86*, 38-47. doi: [10.1016/j.jvb.2014.10.006](https://doi.org/10.1016/j.jvb.2014.10.006)
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
- Little, T. D. (2013). *Longitudinal Structural Equation Mo-*

deling. New York, NY: Guilford.

- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149. doi: [10.1037/1082-989X.1.2.130](https://doi.org/10.1037/1082-989X.1.2.130)
- Preacher, K. J., & Coffman, D. L. (2006). Computing power and minimum sample size for RMSEA [software de cálculo]. Recuperado de <http://quantpsy.org>
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5-26. Recuperado de <https://psycnet.apa.org/record/2012-34046-001>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237. doi: [10.1080/00223891.2015.1089249](https://doi.org/10.1080/00223891.2015.1089249)
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. doi: [10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02)
- Snyder, C. R. (2002). Hope theory: Rainbows in the mind. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, *13*(4), 249-275. doi: [10.1207/S15327965PLI1304_01](https://doi.org/10.1207/S15327965PLI1304_01)
- Snyder, C. R., Feldman, D. B., Taylor, J. D., Schroeder, L., & Adams, V. H. (2000). The roles of hopeful thinking in preventing problems and enhancing strengths. *Applied and Preventive Psychology*, *9*(4), 249-269. doi: [10.1016/S0962-1849\(00\)80003-7](https://doi.org/10.1016/S0962-1849(00)80003-7)
- Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T. ... Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology*, *60*(4), 570-585. doi: [10.1037/0022-3514.60.4.570](https://doi.org/10.1037/0022-3514.60.4.570)
- Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., ... Stahl, K. J. (1997). The development and validation of the Children's Hope Scale. *Journal of Pediatric Psychology*, *22*(3), 399-421. doi: [10.1093/jpepsy/22.3.399](https://doi.org/10.1093/jpepsy/22.3.399)
- Snyder, C. R., Sympson, S. C., Ybasco, F. C., Borders, T. F., Babyak, M. A., & Higgins, R. L. (1996). Development and validation of the State Hope Scale. *Journal of Personality and Social Psychology*, *70*(2), 321-335. doi: [10.1037/0022-3514.70.2.321](https://doi.org/10.1037/0022-3514.70.2.321)
- Stucky, B. D., & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. En S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 183-206). New York, NY: Routledge.
- Yarcheski, A., & Mahon, N. E. (2014). Meta-analyses of predictors of hope in adolescents. *Western Journal of Nursing Research*, *38*(3), 345-368. doi: [10.1177/0193945914559545](https://doi.org/10.1177/0193945914559545)
- Zárate-Torres, R. A., & Acosta-Prado, J. C. (2015). Esperanza y atributos de los colaboradores. *Suma de Negocios*, *6*(14), 178-182. doi: [10.1016/j.sum-neg.2015.07.002](https://doi.org/10.1016/j.sum-neg.2015.07.002)