

ÍTEMS TIPO LIKERT VS. ÍTEMS CONTINUOS: ¿QUÉ OPCIÓN ES MEJOR?

JAIME GARCÍA-FERNÁNDEZ
Universidad de Oviedo.
garciafernandezj@uniovi.es
ORCID iD: <https://orcid.org/0000-0001-7414-8807>

MARCELINO CUESTA
Universidad de Oviedo
mcuesta@uniovi.es
ORCID iD: <https://orcid.org/0000-0002-1060-9536>

ÁLVARO POSTIGO
Universidad de Oviedo
postigoalvaro@uniovi.es
ORCID iD: <https://orcid.org/0000-0003-4228-8965>

ÁLVARO MENÉNDEZ-ALLER
Universidad de Oviedo
alvarezmenendezpsicologo@outlook.com
ORCID iD: <https://orcid.org/0000-0002-4049-5488>

COVADONGA GONZÁLEZ-NUEVO
Universidad de Oviedo
gonzalezvcovadonga@uniovi.es
ORCID iD: <https://orcid.org/0000-0003-0501-8907>

EDUARDO GARCÍA-CUETO
Universidad de Oviedo
egcueto@outlook.com
ORCID iD: <https://orcid.org/0000-0002-1059-117X>

Cómo citar este artículo / Citation: García-Fernández, Jaime, Marcelino Cuesta, Álvaro Postigo, Álvaro Menéndez-Aller, Covadonga González-Nuevo y Eduardo García-Cueto. 2024. Ítems tipo Likert vs. Ítems continuos: ¿Qué opción es mejor?, *Revista Internacional de Sociología* 82 (2): e252. <https://doi.org/10.3989/ris.2024.82.2.M23-04>

Copyright: © 2024 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)

Recibido: 04.11.2022. **Aceptado:** 01.12.2023
Publicado: 11.06.2024

LIKERT-TYPE ITEMS VS CONTINUOUS ITEMS: WHICH OPTION IS BETTER?

RESUMEN

Los ítems tipo Likert son comúnmente utilizados en ciencias sociales. Este estudio comprueba si se pueden mejorar las propiedades psicométricas de los autoinformes variando los formatos de respuesta. Los formatos comparados han sido el tipo Likert de cinco alternativas y el VAS (Visual Analog Scale). Para ello, 584 personas (75.2 % mujeres) con una edad media de 32.9 años ($DT = 13.4$) respondieron a dos cuestionarios *online* de personalidad que fueron aplicados una vez con formato tipo Likert (0-4) y una segunda vez con formato VAS (0-100). La discriminación, fiabilidad y pesos factoriales de las escalas con formato Likert han sido mayores que las del formato VAS. El formato de respuesta no afectó a la estructura factorial. Las escalas con formato VAS dan puntuaciones totales significativamente más altas. Aunque los ítems con formato de respuesta Likert muestran un mejor funcionamiento, estos resultados podrían estar influenciados por la longitud de la línea en el formato VAS.

PALABRAS CLAVE

Likert, VAS, formato de respuesta, escalas psicológicas, test de personalidad.

ABSTRACT

Likert-type items are commonly used in social science research. This study explores if changing the response format of a scale could improve its psychometric properties. The compared response formats were the five-point Likert scale and the VAS (Visual Analog Scale). 584 participants (75.2 % women) with a mean age of 32.9 ages ($SD = 13.4$) answered an online survey VAS with two personality questionnaires that were applied twice, one with Likert items (0-4) and another with VAS items (0-100). Discrimination indexes, reliability, and factor loadings from scales with the Likert response format were higher than the ones found in the VAS scales. The response format did not affect the factorial structure of the scales. The VAS response format scales gave higher scores than the Likert response format scales. Although Likert response format showed better properties, this result may have been influenced by the line length of the VAS scale.

KEYWORDS

Likert, VAS, Response format, Psychological Scales, Personality test.

INTRODUCCIÓN

Los cuestionarios autoinformados son un método de evaluación profusamente utilizado en el ámbito de las ciencias sociales¹. Están formados por una serie de ítems que, conjuntamente, permiten estimar el nivel de una persona en una determinada variable o constructo psicológico. Siguiendo la teoría clásica de los test, una de las formas de conseguir esto es a través de la suma de las puntuaciones en los ítems (Lord, Novick y Birnbaum 1968), obteniendo una puntuación total que ha de guardar una relación monótonica con el nivel en el rasgo medido (Cox 1980).

La asignación de puntuaciones a las respuestas de los sujetos está condicionada por el formato de respuesta de los ítems. El formato Likert es un formato bastante habitual en investigaciones de las ciencias sociales y de la salud. Consiste en preguntar al evaluado por su grado de acuerdo con el enunciado del ítem, ofreciéndole alternativas de respuesta ordenadas, por ejemplo, desde el “completamente en desacuerdo” al “completamente de acuerdo” (Likert 1932). Partiendo del postulado de que parece dudoso que los sujetos entiendan las alternativas como equidistantes (*i. e.*, que las distancias entre las posibles respuestas sean las mismas), las respuestas a los ítems vendrían dadas en una escala ordinal. En consecuencia, teóricamente no sería adecuado realizar operaciones aritméticas con los valores numéricos asignados a las alternativas (Jamieson 2004; Kuzon, Urbanek y McCabe 1996; Stevens 1946), imposibilitando la obtención de una puntuación total en el test a través de la suma de las puntuaciones en los ítems. No obstante, muchos otros autores critican esta perspectiva, alegando que las escalas de medida no limitan las técnicas estadísticas que se pueden aplicar sobre los datos, sino las interpretaciones posteriores de estos análisis (Carifio y Perla 2008; 2007; Pell 2005; Knapp 1990; Lord 1953). Incluso el propio S. S. Stevens, autor de la clasificación de las escalas de medida en nominal, ordinal, intervalo y razón, reconoce que esta práctica “estadísticamente ilegal” puede llevar a resultados fructíferos (Stevens 1946: 679).

Siguiendo esta visión menos “purista”, y dado que los formatos de respuesta tipo Likert presentan multitud de variaciones, es lógico pensar que algunas de estas últimas se acercan más a ser escalas de intervalo. Por ejemplo, algunos autores afirman que eliminar las etiquetas intermedias en los formatos tipo Likert mejora la linealidad de la escala, si bien este es un punto polémico (Kronsick y Fabrigar 1997). El espacio visual entre las opciones de respuesta puede

modificar la proporción de alternativas seleccionadas: si los extremos se acercan a los puntos intermedios, la proporción de respuestas en los extremos baja (Funke y Reips 2006). Incluso el uso de números negativos para identificar opciones de respuesta (p. ej., ir de -2 a 2 en vez de 0 a 5) o el orden —ascendente o descendente— de las alternativas afecta a la respuesta de los sujetos (Christian, Parsons, y Dillman 2009; Schwarz 1999). No obstante, el aspecto más estudiado ha sido el número de alternativas de respuesta adecuadas (Leung 2011; Lozano, García-Cueto y Muñiz 2008; Preston y Colman 2000; Cox 1980). Los resultados más habituales indican que a mayor número de alternativas, mayor precisión en la respuesta del participante, si bien existe un límite (en torno a siete alternativas) a partir del cual este incremento no es estadísticamente significativo (Lozano, García-Cueto y Muñiz 2008; Cox 1980).

Si pudiese aumentarse infinitamente el número de alternativas, se llegaría a una escala continua, en la que parece más probable que el sujeto entienda los puntos de la escala como equidistantes. La idea de ofrecer un formato de respuesta continuo fue introducida a principios del siglo pasado, bajo el término ‘*Graphic Rating Scales*’ (GRS) (Freyd 1923; Hayes y Patterson 1921). Estos formatos consisten en una línea cuyos extremos se identifican con sendas etiquetas (p. ej.: “completamente en desacuerdo” y “completamente de acuerdo”). Las etiquetas intermedias suelen evitarse, ya que se ha demostrado que sesgan la linealidad de la escala (Matejka *et al.* 2016). Las GRS sin etiquetas intermedias suelen identificarse como ‘*Visual Analog Scales*’ (VAS) (Couper *et al.* 2006; Svensson 2000). El sujeto da su respuesta marcando uno de los infinitos puntos que ofrece la línea. La puntuación del ítem se obtiene midiendo la distancia entre uno de los extremos y el punto marcado. Si bien el proceso de puntuación es laborioso en cuestionarios de lápiz y papel, el mismo se vuelve automático en cuestionarios aplicados por ordenador. La implementación informática de las VAS puede hacerse bien replicando la conducta del sujeto al responder con lápiz y papel (*i. e.*, marcar un punto sobre una línea) o utilizando *sliders*, en los cuales la persona da la respuesta arrastrando un punto sobre una línea. Aunque esta diferencia es sutil, se han detectado diferencias entre ambos métodos (Funke 2016).

Así, el uso de un formato de respuesta continuo como el VAS elude el debate acerca del cálculo de puntuaciones totales a partir de ítems con formato de respuesta Likert. En otras palabras, la mejor forma de evitar los problemas derivados de la discretización de las escalas de medida (esto es, el establecimiento de un número limitado de alternativas de respuesta, como ocurre en el formato de respuesta tipo Likert) es no discretizar (Krieg 1999). Así, obviando el componente psicológico del proceso de respuesta, la mejora en la precisión (*i. e.*, la capacidad del participante para situarse en la escala de medida) que permite el

1 Véase, por ejemplo, Stoica y Umbres (2021), Schoene (2019), Huang-Isherwood *et al.* (2022), así como la construcción de la escala EAPDP de Fuentealba y Barriga (2017) o la investigación de Benítez, Van de Vijver y Padilla (2022) sobre el sesgo de este tipo de instrumentos en estudios transculturales.

formato VAS se debería traducir en una mejora en las propiedades psicométricas del cuestionario. Esto se debe a que el aumento del número de alternativas de respuesta debería aumentar la variabilidad de las puntuaciones (Cox 1980), y, en el caso de las VAS, el número de alternativas es infinito. Suponiendo que el participante está respondiendo de forma precisa, con el formato VAS se obtendrían covarianzas mayores entre los ítems que evalúan el mismo rasgo. Esto mejoraría todos los índices basados en la covarianza o correlación, como es el caso de los derivados del análisis factorial o del coeficiente α de Cronbach, algo ya demostrado empíricamente (Preston y Colman 2000). Finalmente, mayor variabilidad en las puntuaciones de los ítems conllevaría mayor variabilidad en las puntuaciones totales de los cuestionarios, mejorando su capacidad discriminativa y los valores correlaciones en las relaciones con otras variables.

Pese a esto, los estudios que comparan empíricamente el formato de respuesta VAS con el Likert no suelen encontrar grandes diferencias entre ellos. Cook *et al.* (2001), encontraron fiabilidades ligeramente mejores en el caso del formato VAS, si bien el tiempo de respuesta fue ligeramente mayor. Briesch *et al.* (2013) y Couper *et al.* (2006) no encontraron diferencias en las varianzas de las respuestas VAS y Likert, en contraste con Guyatt *et al.* (1987), que sí las hallaron. Las diferencias más notables que suelen señalarse son el mayor número de respuestas extremas en los formatos VAS frente a las escalas Likert (Funke y Reips 2006; Hasson y Arnetz 2005). Estas comparaciones no suelen hacerse con escalas que evalúen rasgos de personalidad. Por ejemplo, Funke y Reips (2012) preguntan por los gustos estéticos, Couper *et al.* (2006) por la opinión acerca de la influencia de los genes y ambientes en la conducta, y Roster, Lucia-netti y Albaum (2015) por la sensibilidad a la investigación en el ámbito empresarial. En cuanto a la mayor parte de las investigaciones con escalas VAS, preguntan acerca de cuestiones médicas (Voutilainen *et al.* 2016; Myles y Urquhart 2005; Bijur, Silver y Gallagher 2001; Myles *et al.* 1999). Si bien existe algún estudio que compara escalas de respuesta Likert y continua en cuestionarios de personalidad administrados *online* (Toland *et al.* 2021; Simms *et al.* 2019; Kuhlmann, Dantlgraber y Reips 2017; Hilbert *et al.* 2016), estos no son frecuentes.

Dada la fácil implementación informática de las escalas VAS y los posibles beneficios psicométricos que estas podrían aportar a los cuestionarios de evaluación, nos parece muy interesante seguir recabando datos que permitan confirmar, o no, los beneficios hipotetizados. Desde esta perspectiva, el objetivo del presente trabajo es estudiar la influencia del formato de respuesta (VAS frente a Likert de cinco puntos) en las propiedades psicométricas de escalas de rasgos de personalidad administradas *online*. Para ello, se utilizó un cuestionario conformado por

dos escalas (temeridad e impulsividad), las cuales se administraron en ambos formatos (Likert y VAS) entre febrero y mayo de 2022.

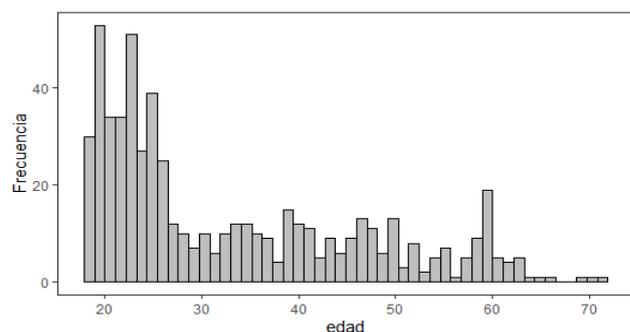
A continuación, en el apartado de método se indica la muestra recogida, el procedimiento de aplicación, los instrumentos utilizados y el plan de análisis de datos. Posteriormente, se presentan los resultados derivados de la investigación. Finalmente, en el apartado de discusión estos resultados se ponen en relación con el objetivo anteriormente mencionado y se describen algunas limitaciones del presente estudio.

METODOLOGÍA: PARTICIPANTES, PROCEDIMIENTO, INSTRUMENTOS, ANÁLISIS DE DATOS

Participantes

El cuestionario fue respondido por 695 personas, de las cuales 110 fueron excluidas por fallar más de una pregunta de control atencional (ver Instrumentos). Asimismo, un participante fue excluido por responder 2 años a la pregunta de edad. La muestra final estuvo conformada por 584 participantes, siendo el 75.2 % mujeres. La edad media fue de 32.9 años (DT = 13.4). En la figura 1, puede consultarse el histograma de la edad. El 52.91 % de los participantes contaban con estudios superiores (i. e., universitarios, FP de grado superior), el 43.67 % con estudios posobligatorios (i. e., bachiller, FP de grado medio) y el 3.42 % con estudios básicos u obligatorios.

Figura 1
Histograma de la edad de los participantes



Se utilizó un muestreo de conveniencia, difundiendo el enlace del cuestionario por distintas redes sociales (i. e., Whatsapp, Facebook, Instagram, LinkedIn) y solicitando a los participantes su difusión. Atendiendo al código deontológico (Consejo General de la Psicología de España 2015) y a la Ley de Protección de Datos (2018), la participación fue anónima y voluntaria. Todos los participantes aceptaron el consentimiento informado previamente a la cumplimentación del cuestionario. Las personas no recibieron ningún tipo de recompensa económica por participar.

PROCEDIMIENTO

Las escalas aplicadas formaban parte de una investigación más amplia sobre rasgos socialmente aversivos en que los participantes debían responder a un cuestionario *online*, accesible desde móvil, tableta u ordenador. Este cuestionario fue subido a una página web diseñada *ad hoc* por los autores. Los ítems con formato de respuesta tipo Likert consistían en cinco botones de opción, mientras que los formatos VAS eran *sliders* HTML de 150 píxeles de longitud. Se fijó esta longitud puesto que la misma permitía responder cómodamente tanto desde un ordenador como desde dispositivos con menor tamaño de pantalla (figura 2).

Tras contestar a las preguntas referentes a variables sociodemográficas (edad, sexo, nivel educativo), el participante comenzaba respondiendo los ítems tipo Likert correspondientes a las escalas cortoplacismo y temeridad, los cuales estaban mezclados con las escalas de honestidad-humildad y amabilidad del cuestionario de personalidad HEXACO (Roncero, Fornés y Belloch 2013), la versión española del cuestionario de psicopatía SRP-III (Gómez-Leal et al. 2021) y la escala de triada oscura DD (Pineda, Sandín y Muris 2020). Posteriormente, respondía a las preguntas en formato VAS de las escalas cortoplacismo y temeridad, así como de otras 2 escalas (manipulación e insensibilidad) que forman parte de un cuestionario de evaluación psicológica pendiente de publicación por los autores del manuscrito. El cuestionario total estaba formado por 79 ítems con formato de respuesta Likert y 26 ítems con formato de respuesta VAS

Figura 2.

Ejemplo de formato de respuesta Likert y VAS

6. Hay que vivir el presente como si no hubiera un mañana.



42. Hay que vivir el presente como si no hubiera un mañana.



INSTRUMENTOS

Las tres escalas pueden definirse de la forma siguiente²:

2 Las dos primeras escalas (cortoplacismo y temeridad) forman parte de un cuestionario desarrollado por los autores del manuscrito, el cual está pendiente de publicación. Las fiabilidades (estimadas con el coeficiente α) para la versión Likert y la VAS de sendas escalas puede consultarse en la parte dedicada a los resultados de investigación.

- *Escala de cortoplacismo*. Se trata de una escala unidimensional formada por seis ítems, los cuales evalúan tendencias cortoplacistas (*i. e.*, preferencia por reforzadores cercanos frente a demorados) utilizando un formato de respuesta tipo Likert de cinco opciones. Sus ítems pueden consultarse en la tabla 1.
- *Escala de temeridad*. Esta escala unidimensional está formada por cinco ítems, los cuales evalúan el gusto por la práctica de actividades de riesgo. También utiliza un formato de respuesta tipo Likert de cinco opciones. Sus ítems pueden consultarse en la tabla 1.

Tabla 1.

Ítems de la escala de cortoplacismo y temeridad

Cortoplacismo
1. Sigo mis impulsos sin intentar controlarlos.
2. Prefiero lo inmediato.
3. La mayoría de mis intereses son a corto plazo.
4. Pensar en el mañana es una pérdida de tiempo.
5. Hay que vivir el presente como si no hubiera un mañana.
6. Prefiero estar contento ahora que buscar la estabilidad a largo plazo.
Temeridad
1. Cuando me juego la vida es cuando más vivo me siento.
2. Me encantan las emociones fuertes.
3. Me gusta hacer cosas peligrosas.
4. Me encantaría tirarme en paracaídas.
5. Practico deportes de alto riesgo.

- *Escala de control atencional*. Esta escala está formada por 10 ítems que piden al participante responder una opción determinada (p. ej.: “Por favor, seleccione completamente de acuerdo”). De esta forma, se espera controlar si la persona está leyendo los contenidos de los ítems o, por el contrario, está respondiendo al azar. Si el participante responde equivocadamente a más de una de estas preguntas, la persona es eliminada del estudio, ya que se entiende que no está respondiendo de forma rigurosa a los ítems.

ANÁLISIS DE DATOS

Inicialmente, para poder comparar ambos formatos de respuesta, se transformaron linealmente las puntuaciones de los formatos VAS (0-100) para que estuviesen entre los límites de las respuestas Likert (0-4), permitiendo así su comparación. Con el objetivo de analizar si existen diferencias entre los dos formatos a nivel de ítem, se calcularon sus medias, desviaciones típicas e índices de discriminación. Además, se representaron sus histogramas, superponiendo los de los ítems con formato Likert sobre los de formato VAS.

Para analizar si las covarianzas entre los ítems de las escalas VAS fueron superiores a las de las escalas Likert, se calculó la matriz de correlaciones entre los ítems de cada escala.

Posteriormente, para comprobar si existían diferencias entre las puntuaciones totales de las escalas Likert y VAS, se compararon sus medias con la prueba *t* de Student —aportando la *d* de Cohen como estimador del tamaño del efecto (Cohen 1992)— y sus varianzas con la prueba de Brown-Forsythe.

A continuación, se llevó a cabo un análisis factorial exploratorio con cada escala, utilizando correlaciones de Pearson y extrayendo el número de dimensiones recomendadas por el análisis paralelo. Como método de extracción, se utilizó el de mínimos cuadrados no ponderados robustos. El ajuste de la estructura factorial se comprobó con el porcentaje de varianza explicada y los estadísticos CFI (*Comparative Fit Index* o índice de ajuste comparativo) y RMSEA (*Root Mean Square Error of Aproximation* o raíz cuadrada de la media de los errores cuadráticos de aproximación). Para comprobar si el formato de respuesta afectó a la estructura factorial de cada escala, se compararon sus

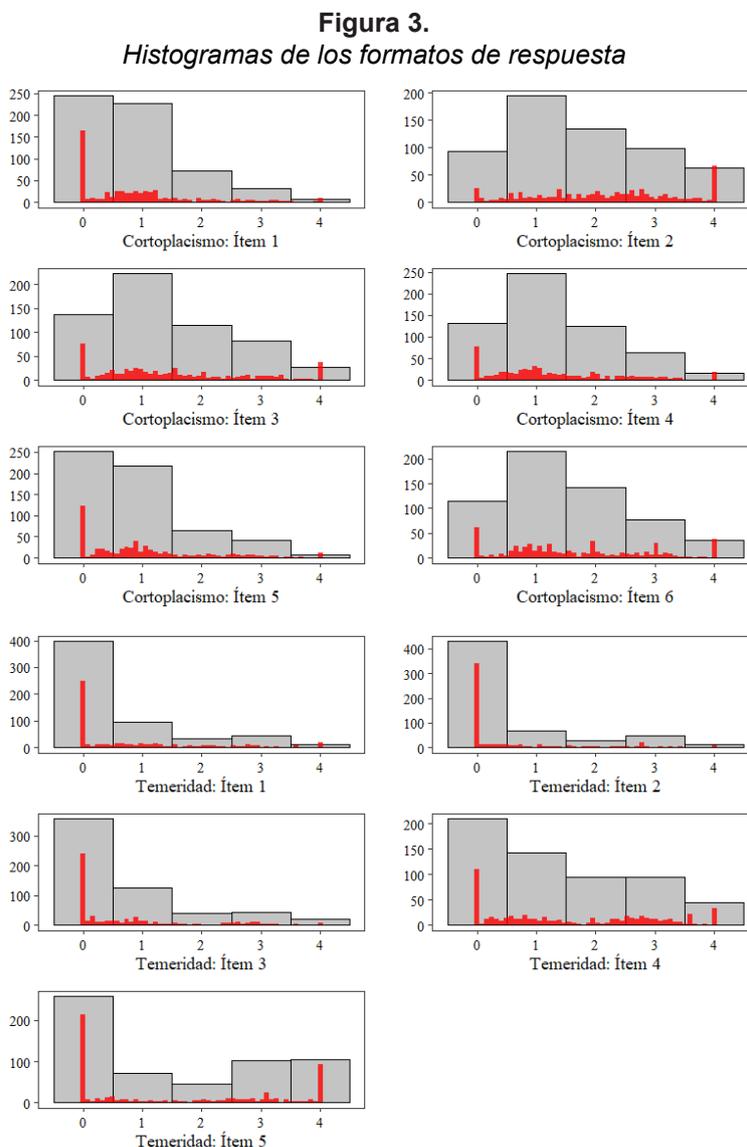
pesos factoriales con el coeficiente de congruencia de Wrigley y Neuhaus (García-Cueto 1994) y sus porcentajes de varianza explicada.

Posteriormente, con la finalidad de observar si el formato de respuesta VAS mejora la fiabilidad de la escala, se estimó el coeficiente alpha de Cronbach (α) de cada una de ellas, comparándolos entre sí con el estadístico *t* (Feldt 1980).

Todos los análisis estadísticos se han realizado en R, versión 4.1.1 (R Core Team 2022), con excepción de los análisis factoriales, realizados con el programa Factor, versión 12.01.01 (Lorenzo-Seva y Ferrando 2006). Los resultados y sintaxis de análisis están accesibles en el material complementario.

RESULTADOS

Los histogramas de los formatos Likert y VAS pueden consultarse en la figura 3. Para el cortopla-



Nota. En gris, los histogramas de los formatos de respuesta Likert. En rojo, los histogramas de los formatos de respuesta VAS

cismo, las respuestas VAS se distribuyen por toda la escala, si bien las respuestas más frecuentes se dan en los extremos. Esta tendencia no se observa en los ítems Likert, en los cuales la alternativa cuatro (“completamente de acuerdo”) es siempre la menos seleccionada. En el caso de la temeridad, se observa un efecto suelo claro (i. e., gran proporción de respuestas en la alternativa más baja de la escala) en ambos formatos de respuesta.

Las bajas frecuencias de respuestas en el extremo derecho de los ítems de las escalas Likert frente a las encontradas en las escalas VAS explican por qué las respuestas VAS cuentan con una media mayor (tabla 2). Las desviaciones típicas también son ligeramente mayores en el caso de los ítems de las escalas VAS.

En cuanto a los índices de discriminación, los formatos de respuesta Likert han obtenido mejores índices que los formatos de respuesta VAS (tabla 2). Las correlaciones entre ítems fueron mayores en las escalas Likert frente a las VAS (tabla 3).

Respecto a las puntuaciones totales, se encontraron diferencias estadísticamente significativas entre las escalas con formato VAS y las Likert, siendo superiores las primeras (cortoplacismo: $t = 10.671$, $p < .001$, $d = 0.257$; temeridad: $t = -11.121$, $p < .001$, $d = 0.177$). No obstante, no se han observado diferencias estadísticamente significativas entre las varianzas de las puntuaciones totales (cortoplacismo: $F(1, 1166) = 0.842$, $p = .718$; temeridad: $F(1, 1166) = 0.750$, $p = .774$).

El porcentaje de varianza explicada, así como el RMSEA y el CFI de cada escala, pueden consultarse

Tabla 2.
Estadísticos descriptivos de los ítems

Ítems	Medias		D. típicas		I. Discriminación	
	Likert	VAS	Likert	VAS	Likert	VAS
Cortoplacismo						
Ítem 1	0.853	0.940	0.927	0.980	.423	0.319
Ítem 2	1.729	2.164	1.222	1.173	.424	0.384
Ítem 3	1.380	1.513	1.122	1.187	.651	0.493
Ítem 4	1.289	1.328	1.021	1.062	.577	0.526
Ítem 5	0.861	1.103	0.962	1.027	.384	0.372
Ítem 6	1.497	1.670	1.130	1.167	.682	0.580
Total	7.609	8.717	4.398	4.236		
Temeridad						
Ítem 1	0.591	0.891	1.029	1.135	.586	0.511
Ítem 2	0.533	0.626	1.037	1.058	.537	0.469
Ítem 3	0.699	0.812	1.084	1.077	.780	0.739
Ítem 4	1.349	1.629	1.313	1.337	.771	0.683
Ítem 5	1.529	1.583	1.606	1.608	.600	0.563
Total	4.700	5.540	4.793	4.682		

Nota. D. típicas = desviaciones típicas. I. Discriminación = índices de discriminación

en la tabla 4. Si bien los ajustes fueron similares, las escalas Likert cuentan con un mayor porcentaje de varianza explicada.

Los pesos factoriales para cada escala pueden consultarse en la tabla 5. Los índices de congruencia (cortoplacismo: .995; temeridad: .999) indican que la estructura factorial del cuestionario no se ha visto afectada por el formato de respuesta.

Tabla 3.
Correlaciones entre los ítems

Cortoplacismo						
Likert \ VAS	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6
Ítem 1	-	.248	.222	.190	.135	.277
Ítem 2	.249	-	.235	.270	.209	.323
Ítem 3	.346	.330	-	.458	.237	.441
Ítem 4	.369	.318	.579	-	.310	.444
Ítem 5	.200	.247	.292	.222	-	.348
Ítem 6	.353	.373	.653	.492	.417	-
Temeridad						
Likert \ VAS	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	
Ítem 1	-	.314	.612	.435	.292	
Ítem 2	.359	-	.419	.424	.339	
Ítem 3	.674	.492	-	.614	.541	
Ítem 4	.568	.504	.718	-	.559	
Ítem 5	.351	.416	.557	.605	-	

Nota. Por encima de la diagonal de cada matriz, correlaciones entre los ítems de la escala VAS. Por debajo de la diagonal, correlaciones entre los ítems de la escala Likert

Tabla 4.
Ajuste de las soluciones factoriales

Escala	% var,	RMSR	CFI
Cortoplacismo			
Likert	47.91	.041	.99
Continua	41.52	.036	.99
Temeridad			
Likert	62.50	.046	.98
Continua	56.95	.052	.97

Nota. % var, = porcentaje de varianza explicada. RMSR = raíz cuadrada de la media de errores cuadráticos de aproximación
CFI = índice de ajuste comparativo

Tabla 5.
Pesos factoriales de los ítems

Ítems	Pesos factoriales	
	Likert	VAS
Cortoplacismo		
Ítem 1	.475	.367
Ítem 2	.469	.445
Ítem 3	.787	.615
Ítem 4	.671	.650
Ítem 5	.428	.448
Ítem 6	.804	.720
Temeridad		
Ítem 1	.669	.604
Ítem 2	.578	.520
Ítem 3	.884	.862
Ítem 4	.857	.768
Ítem 5	.647	.634

Respecto a la fiabilidad, las estimaciones a través del coeficiente α indican una consistencia interna mayor en las escalas con formato Likert frente a las VAS, diferencias estadísticamente significativas según la t de Feldt (cortoplacismo: $\alpha_{\text{Likert}} = 0.775$, $\alpha_{\text{VAS}} = 0.713$, $t = 5.278$, $p < .001$; temeridad: $\alpha_{\text{Likert}} = 0.836$, $\alpha_{\text{VAS}} = 0.797$, $t = 6.745$, $p < .001$).

DISCUSIÓN

El objetivo de esta investigación fue comprobar si las escalas de respuesta VAS mejoraban las propiedades psicométricas de dos escalas de personalidad administradas *online* cuyos formatos de respuesta originales eran tipo Likert de cinco alternativas. Partiendo del hecho de que, si bien existen investigaciones que comparan estos formatos de respuesta (Voutilainen *et al.* 2016; Roster, Lucianetti y Albaum 2015; Briesch *et al.* 2013; Funke y Reips 2012; Reips y Funke 2008; Couper *et al.* 2006; Hasson y Arnetz 2005; Cook *et al.* 2001; Guyatt *et al.* 1987), las investigaciones que utilizan

cuestionarios de rasgos de personalidad no suelen ser muy frecuentes (Toland *et al.* 2021; Simms *et al.* 2019; Kuhlmann, Dantlgraber y Reips 2017; Hilbert *et al.* 2016), los resultados indican diferencias en las puntuaciones totales, índices de discriminación y estimaciones de fiabilidad; no obstante, el formato de respuesta no afectó a la varianza de las puntuaciones totales ni a la estructura factorial.

Una inspección visual de los histogramas muestra que en los ítems con formato de respuesta Likert hay menos respuestas extremas en la parte derecha de la escala comparada que en las encontradas en el formato VAS. Este efecto se conoce como '*end-aversion bias*' (Funke y Reips 2006; Hasson y Arnetz 2005). No obstante, esta tendencia suele ser bilateral, es decir, las personas suelen responder en los puntos intermedios de la escala, evitando sus extremos. En este caso, la aversión únicamente se da en el extremo derecho de los formatos Likert. Este fenómeno puede explicar por qué la media de los ítems VAS es mayor que la de los ítems Likert, pues la presencia de puntuaciones extremas en la alternativa de mayor valor empuja la media hacia arriba. La presencia de puntuaciones extremas en la escala VAS podría explicarse por una mayor facilidad de respuesta en los extremos (*i. e.*, arrastrar el *slider* a un extremo frente a ajustar un valor bajo o alto).

Pese a que no se han observado diferencias estadísticamente significativas entre las medias o varianzas de las puntuaciones totales, sí se han observado efectos en estimadores derivados de la varianza, como es el caso de α . Revisando la tabla 2, puede observarse que las desviaciones típicas de los ítems son ligeramente más pequeñas en el caso de los ítems con formato Likert. Este hecho puede tener un efecto notable en α (manteniendo constante la varianza de la puntuación total), explicando así el hecho de haber encontrado evidencias de mayor consistencia interna en las escalas Likert. Una mayor consistencia interna implica correlaciones más altas entre los ítems de la escala y mayores índices de discriminación, lo que esta investigación ha probado empíricamente.

Las diferencias de fiabilidad y discriminación indican que los participantes son más precisos con el formato de respuesta Likert. Esta mayor precisión puede deberse a la dificultad que algunas personas pueden tener para responder utilizando *sliders*, especialmente en dispositivos táctiles (Buskirk 2015). Otro motivo puede ser la poca longitud de la línea de los formatos VAS, cuyo tamaño se fijó para poder responder cómodamente desde dispositivos pequeños (150 píxeles equivalen a 5,3 centímetros en una resolución de 720p., la más habitual en pantallas electrónicas). El hecho de haber demostrado que la longitud de la línea en escalas VAS influye en la precisión del sujeto (Reips y Funke 2008) nos lleva a

concluir que el peor rendimiento de las escalas VAS se debe probablemente a esta cuestión.

Finalmente, aunque las estructuras factoriales de las escalas no se han visto afectadas por el formato de respuesta, el porcentaje de varianza explicada ha sido menor en el caso de las escalas continuas. Esto último se debe al menor valor de los pesos factoriales, derivados de la disminución en las covarianzas entre los ítems. Dicho de otro modo, la estructura factorial es la misma (lo cual no es sorprendente, teniendo en cuenta que se trata de los mismos ítems), pero está menos “definida” debido a la menor precisión del formato de respuesta VAS.

Esta investigación muestra que el efecto de un formato de respuesta sobre las propiedades psicométricas de un test depende, fundamentalmente, del cambio que este produzca en las correlaciones entre los ítems (*i. e.*, la consistencia interna). Uno de los factores que influirá en la consistencia interna es la precisión del sujeto al utilizar la escala de respuesta. Si bien se ha hipotetizado que las escalas VAS permiten una mayor precisión (Voutilainen *et al.* 2016), otros estudios empíricos concluyen que las propiedades psicométricas prácticamente no varían entre ambos formatos de respuesta (Briesch *et al.* 2013; Couper *et al.* 2006; Cook *et al.* 2001). De hecho, algunas investigaciones reivindican el uso de los formatos Likert frente a los VAS, apoyándose en la sensación de que las personas parecen entenderlos mejor (Hilbert *et al.* 2016; van Laerhoven, van der Zaag-Loonen, y Derkx 2004; Guyatt *et al.* 1987) y que, en ocasiones, sus propiedades psicométricas son mejores (Simms *et al.* 2019).

Aunque esta investigación ahonda en la línea de esta última afirmación, existen una serie de limitaciones que han de tenerse en cuenta. La primera es que no se han probado distintas longitudes de línea en los formatos VAS, pudiendo darse el caso de que, con líneas más largas, la precisión del sujeto mejor. La segunda es que tampoco se ha registrado el dispositivo desde el que ha respondido el participante, lo cual impide comprobar si existen diferencias al responder desde un teléfono, tableta u ordenador. Finalmente, el haber realizado un muestreo de conveniencia puede haber introducido ciertos sesgos (véanse, por ejemplo, las diferencias de participación entre hombres y mujeres), lo cual limita la generalización de los resultados. Cuestiones interesantes para futuras investigaciones podrían ser la inclusión de cuestionarios de lápiz y papel y el uso de cuestionarios que evalúen otros tipos de rasgos.

AGRADECIMIENTOS:

Queremos expresar nuestro agradecimiento a todos los participantes involucrados en este estudio por su tiempo y cooperación. Sin su colaboración altruista, no hubiese sido posible desarrollar esta investigación.

FUENTES DE FINANCIACIÓN:

Esta investigación ha sido subvencionada por el Ministerio de Ciencia, Innovación y Universidades del Gobierno de España en el marco de una ayuda para la Formación del Profesorado Universitario (FPU-21-01121).

DECLARACIÓN DE CONFLICTOS DE INTERESES

Los autores de este artículo declaran no tener conflictos de intereses financieros, profesionales o personales que pudieran haber influido de manera inapropiada en este trabajo.

DECLARACIÓN DE CONTRIBUCIÓN DE AUTORÍA

Jaime García-Fernández: conceptualización, análisis formal, investigación, metodología, administración de proyecto, redacción – borrador original, redacción – revisión.

Marcelino Cuesta: conceptualización, análisis formal, investigación, metodología, administración de proyecto, redacción – borrador original, redacción – revisión.

Álvaro Postigo: conceptualización, análisis formal, metodología, administración de proyecto, redacción – borrador original, redacción – revisión.

Álvaro Menéndez-Aller: conceptualización, análisis formal, metodología, administración de proyecto, redacción – borrador original, redacción – revisión.

Covadonga González-Nuevo: conceptualización, análisis formal, metodología, administración de proyecto, redacción – borrador original, redacción – revisión.

Eduardo García-Cueto: conceptualización, análisis formal, investigación, metodología, administración de proyecto, redacción – revisión.

REFERENCIAS

- Benítez, Isabel, Fons Van De Vijver y José Luis Padilla. 2022. “A Mixed Methods Approach to the Analysis of Bias in Cross-Cultural Studies”. *Sociological Methods & Research* 51(1): 237-70. doi: 10.1177/0049124119852390.
- Bijur, Polly E., Wendy Silver y John E. Gallagher. 2001. “Reliability of the Visual Analog Scale for Measurement of Acute Pain”. *Academic Emergency Medicine* 8(12): 1153-57. doi: 10.1111/j.1553-2712.2001.tb01132.x.
- Briesch, Amy M., Stephen P. Kilgus, Sandra M. Chafouleas, T. Chris Riley-Tillman y Theodore J. Christ. 2013. “The Influence of Alternative Scale Formats on the Generalizability of Data Obtained From Direct Behavior Rating Single-Item Scales (DBR-SIS)”. *Assessment for Effective Intervention* 38(2): 127-33. doi: 10.1177/1534508412441966.
- Buskirk, Trent D. 2015. “Are Sliders Too Slick for Surveys? An Experiment Comparing Slider and Radio Button Scales for Smartphone, Tablet and Computer Based Surveys”. *Methods, Data, Analyses* 9(2): 229-60. doi: 10.12758/MDA.2015.013.

- Carifio, James y Rocco Perla. 2008. "Resolving the 50-Year Debate Around Using and Misusing Likert Scales". *Medical Education* 42(12): 1150-52. doi: 10.1111/j.1365-2923.2008.03172.x.
- Carifio, James y Rocco J. Perla. 2007. "Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes". *Journal of Social Sciences* 3(3): 106-16. doi: 10.3844/jssp.2007.106.116.
- Christian, Leah Melani, Nicholas L. Parsons y Don A. Dillman. 2009. "Designing Scalar Questions for Web Surveys". *Sociological Methods & Research* 37(3): 393-425. doi: 10.1177/0049124108330004.
- Cohen, Jacob. 1992. "A power primer". *Psychological Bulletin* 112(1): 155-59. doi: 10.1037/0033-2909.112.1.155.
- Consejo General de la Psicología de España. 2015. *Código Deontológico del Psicólogo*. Consultado el 8 de febrero de 2024 (<https://www.cop.es/index.php?page=CodigoDeontologico>).
- Cook, Colleen, Fred Heath, Russel L. Thompson y Bruce Thompson. 2001. "Score Reliability in Webor Internet-Based Surveys: Unnumbered Graphic Rating Scales versus Likert-Type Scales". *Educational and Psychological Measurement* 61(4): 697-706. doi: 10.1177/00131640121971356.
- Couper, Mick P., Roger Tourangeau, Frederick G. Conrad y Eleanor Singer. 2006. "Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment". *Social Science Computer Review* 24(2): 227-45. doi: 10.1177/0894439305281503.
- Cox, Eli P. 1980. "The Optimal Number of Response Alternatives for a Scale: A Review". *Journal of Marketing Research* 17(4): 407-22. doi: 10.2307/3150495.
- Feldt, Leonard S. 1980. "A Test of the Hypothesis That Cronbach's Alpha Reliability Coefficient Is the Same for Two Tests Administered to the Same Sample". *Psychometrika* 45(1): 99-105. doi: 10.1007/BF02293600.
- Freyd, Max. 1923. "The Graphic Rating Scale". *Journal of Educational Psychology* 14(2): 83-102. doi: 10.1037/h0074329.
- Fuentealba Carrasco, Pablo Javier y Omar A. Barriga. 2017. "Primera aproximación a la construcción de una Escala de Adhesión a los Principios del Derecho Penal (EAPDP). Estudio piloto en el Conurbano del Gran Concepción, Chile". *Revista Internacional de Sociología* 75(2): 060. doi: 10.3989/ris.2017.75.2.15.103.
- Funke, Frederik. 2016. "A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales". *Social Science Computer Review* 34(2): 244-54. doi: 10.1177/0894439315575477.
- Funke, Frederik y Ulf-Dietrich Reips. 2006. "Visual Analogue Scales in Online Surveys: Non-Linear Data Categorization by Transformation with Reduced Extremes". Presentado en General Online Research Conference, Bielefeld.
- Funke, Frederik y Ulf-Dietrich Reips. 2012. "Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales". *Field Methods* 24(3): 310-27. doi: 10.1177/1525822X12444061.
- García-Cueto, Eduardo 1994. "Coeficiente de Congruencia". *Psicothema* 6(3): 465-68.
- Gómez-Leal, Raquel, Alberto Megías-Robles, María José Gutiérrez-Cobo, Rosario Cabello, Enrique G. Fernández-Abascal y Pablo Fernández-Berrocal. 2021. "Spanish Adaptation and Validation of the 34-Item Self-Report Psychopathy Scale (SRP)". *Journal of Personality Disorders* 35(2): 217-35. doi: 10.1521/pedi_2019_33_434.
- Guyatt, Gordon H., Marie Townsend, Leslie B. Berman y Jana L. Keller. 1987. "A Comparison of Likert and Visual Analogue Scales for Measuring Change in Function". *Journal of Chronic Diseases* 40(12): 1129-33. doi: 10.1016/0021-9681(87)90080-4.
- Hasson, Dan y Bengt B. Arnetz. 2005. "Validation and Findings Comparing VAS Vs. Likert Scales for Psychosocial Measurements". *International Electronic Journal of Health Education* 8: 178-92.
- Hayes, Marry H. S. y Donald G. Patterson. 1921. "Experimental Development of the Graphic Rating Method". *Psychological Bulletin* 18(2): 98-99. doi: 10.1037/h0064147.
- Hilbert, Sven, Helmut Küchenhoff, Nina Sarubin, Tristan T. Nakagawa y Markus Bühner. 2016. "The Influence of the Response Format in a Personality Questionnaire: An Analysis of a Dichotomous, a Likert-Type, and a Visual Analogue Scale". *Testing, Psychometrics, Methodology in Applied Psychology* 23(1): 3-24. doi: 10.4473/TPM23.1.1.
- Huang-Isherwood, Ke M., Steffie S. Y. Kim, Dmitri Williams y Alexander J. Bisberg. 2022. "Las mujeres sostienen (más de) la mitad del cielo: examinando las motivaciones, los comportamientos y el capital social en un juego multijugador popular entre las jugadoras". *Revista Internacional de Sociología* 80(4): e219. doi: 10.3989/ris.2022.80.4.M22-003.
- Jamieson, Susan. 2004. "Likert Scales: How to (Ab)Use Them". *Medical Education* 38(12): 1217-18. doi: 10.1111/j.1365-2929.2004.02012.x.
- Knapp, Thomas R. 1990. "Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy". *Nursing Research* 39(2): 121-23. doi: 10.1097/00006199-199003000-00019.
- Krieg, Edward F. 1999. "Biases Induced by Coarse Measurement Scales". *Educational and Psychological Measurement* 59(5): 749-66. doi: 10.1177/00131649921970125.
- Kronsick, Jon A. y Leandre R. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys" Pp. 141-164 . en *Survey Measurement and Process Quality, Wiley Series in Probability and Statistics*, editado por Lars Lyberg, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz, y Dennis Trewin. New York: John Wiley.
- Kuhlmann, Tim, Michael Dantlgraber y Ulf-Dietrich Reips. 2017. "Investigating Measurement Equivalence of Visual Analogue Scales and Likert-Type Scales in Internet-Based Personality Questionnaires". *Behavior Research Methods* 49(6): 2173-81. Doi: 10.3758/s13428-016-0850-x.
- Kuzon, William M., Melanie G. Urbanchek y Steven McCabe. 1996. "The Seven Deadly Sins of Statistical Analysis". *Annals of Plastic Surgery* 37(3): 265-72. Doi: 10.1097/00006637-199609000-00006.
- Van Laerhoven, H., HJ van der Zaag-Loonen y Bhf Derkx. 2004. "A Comparison of Likert Scale and Visual Analogue Scales as Response Options in Children's Questionnaires". *Acta Paediatrica* 93(6): 830-35. Doi: 10.1111/j.1651-2227.2004.tb03026.x.
- Leung, Shing-On. 2011. "A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales". *Journal of Social Service Research* 37(4): 412-21. Doi: 10.1080/01488376.2011.580697.

- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. 2018. *Boletín Oficial del Estado*, 294: 119788-119857 <https://www.boe.es/eli/es/lo/2018/12/05/3>.
- Likert, Rensis. 1932. "A Technique for the Measurement of Attitudes". *Archives of Psychology* 22(140): 5-55.
- Lord, Frederic M. 1953. "On the Statistical Treatment of Football Numbers". *American Psychologist* 8(12): 750-51. Doi: 10.1037/h0063675.
- Lord, Frederic M., Melvin R. Novick y Allan Birnbaum. 1968. *Statistical Theories of Mental Test Scores*. Nachdr. Der Ausg. Reading, Mass. [u.a.]. Charlotte, NC: Information Age Publ.
- Lorenzo-Seva, Urbano y Pere J. Ferrando. 2006. "Factor: A Computer Program to Fit the Exploratory Factor Analysis Model". *Behavior Research Methods* 38(1): 88-91. Doi: 10.3758/BF03192753.
- Lozano, Luis M., Eduardo García-Cueto y José Muñiz. 2008. "Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales". *Methodology* 4(2): 73-79. Doi: 10.1027/1614-2241.4.2.73.
- Matejka, Justin, Michael Glueck, Tovi Grossman y George Fitzmaurice. 2016. "The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales". Pp. 5421-32 en *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, editado por Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, Juan P. Hourcade, Loren Terveen, Scooter Morris. San Jose California USA: ACM.
- Myles, Paul S. y Nicole Urquhart. 2005. "The Linearity of the Visual Analogue Scale in Patients with Severe Acute Pain". *Anaesthesia and Intensive Care* 33(1): 54-58. doi: 10.1177/0310057X0503300108.
- Myles, Paul S., Sally Troedel, Michael Boquest y Mark Reeves. 1999. "The Pain Visual Analog Scale: Is It Linear or Nonlinear?". *Anesthesia & Analgesia* 89(6): 1517-20. doi: 10.1097/0000539-199912000-00038.
- Pell, Godfrey. 2005. "Use and Misuse of Likert Scales". *Medical Education* 39(9): 970-970. doi: 10.1111/j.1365-2929.2005.02237.x.
- Pineda, David, Bonifacio Sandín y Peter Muris. 2020. "Psychometrics properties of the Spanish version of two Dark Triad scales: The Dirty Dozen and the Short Dark Triad". *Current Psychology* 39(5): 1873-81. doi: 10.1007/s12144-018-9888-5.
- Preston, Carolyn C. y Andrew M. Colman. 2000. "Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences". *Acta Psychologica* 104(1): 1-15. doi: 10.1016/S0001-6918(99)00050-5.
- R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Reips, Ulf-Dietrich y Frederik Funke. 2008. "Interval-Level Measurement with Visual Analogue Scales in Internet-Based Research: VAS Generator". *Behavior Research Methods* 40(3): 699-704. doi: 10.3758/BRM.40.3.699.
- Roster, Catherine A., Lorenzo Lucianetti y Gerald Albaum. 2015. "Exploring Slider vs. Categorical Response Formats in Web-Based Surveys". *Journal of Research Practice* 11(1): preprint 1.
- Roncero, María, Gertrudis Fornés y Amparo Belloch. 2013. "Hexaco: Una nueva aproximación a la evaluación de la personalidad en español". *Revista Argentina de Clínica Psicológica* XXII(3): 205-217.
- Schoene, Matthew. 2019. "European Disintegration? Euroscepticism and Europe's Rural/Urban Divide". *European Politics and Society* 20(3): 348-64. doi: 10.1080/23745118.2018.1542768.
- Schwarz, Norbert. 1999. "Self-Reports: How the Questions Shape the Answers". *American Psychologist* 54(2): 93-105. doi: 10.1037/0003-066X.54.2.93.
- Simms, Leonard J., Kerry Zelazny, Trevor F. Williams y Lee Bernstein. 2019. "Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data". *Psychological Assessment* 31(4):557-66. doi: 10.1037/pas0000648.
- Stevens, Stanley S. 1946. "On the Theory of Scales of Measurement". *Science* 103(2684): 677-80. doi: 10.1126/science.103.2684.677.
- Stoica, Cătălin Augustin y Radu Umbreș. 2021. "Suspicious Minds in Times of Crisis: Determinants of Romanians' Beliefs in COVID-19 Conspiracy Theories". *European Societies* 23(sup1): S246-61. doi: 10.1080/14616696.2020.1823450.
- Svensson, Elisabeth. 2000. "Comparison of the Quality of Assessments Using Continuous and Discrete Ordinal Rating Scales". *Biometrical Journal* 42(4): 417-34. doi: 10.1002/1521-4036(200008)42:4<417::AID-BIMJ417>3.0.CO;2-Z.
- Toland, Michael D., Caihong Li, Jonathan Kodet y Robert J. Reese. 2021. "Psychometric Properties of the Outcome Rating Scale: An Item Response Theory Analysis". *Measurement and Evaluation in Counseling and Development* 54(2): 90-105. doi: 10.1080/07481756.2020.1745647.
- Voutilainen, Ari, Taina Pitkääho, Tarja Kvist y Katri Vehviläinen-Julkunen. 2016. "How to Ask About Patient Satisfaction? The Visual Analogue Scale Is Less Vulnerable to Confounding Factors and Ceiling Effect Than a Symmetric Likert Scale". *Journal of Advanced Nursing* 72(4): 946-57. doi: 10.1111/jan.12875.