

Investigación empírica y análisis teórico

## Efecto del número de opciones de respuesta en las propiedades psicométricas de cuatro escalas psicosociales

### Effect of number of response options in psychometric properties of four psychosocial scales

Calleja, Nazira<sup>1\*</sup>; Reskala Sánchez, Félix Javier<sup>1</sup>; Rivera-Fong, Liliana<sup>1</sup> yBuenrostro Mercado, Diana<sup>1</sup>

---

**Resumen:**

Las escalas psicosociales son utilizadas profusamente en estudios de muy diversos campos. Sin embargo, la investigación psicométrica no ha sido concluyente respecto del número de opciones de respuesta que resulta óptimo en la medición de un constructo. El objetivo del presente estudio fue comparar las propiedades psicométricas de cuatro escalas psicosociales empleando diferente número de opciones de respuesta. Los participantes (757 estudiantes universitarios) fueron asignados a uno de tres grupos independientes para responder a una batería integrada por cuatro escalas cortas, que evaluaban gratitud, optimismo, bienestar subjetivo y soledad, con cinco, seis o siete opciones de respuesta. Se encontró una asociación positiva entre el número de opciones y la calidad psicométrica de las escalas. La mayor variabilidad (dispersión) se observó en las versiones de seis y siete opciones; asimismo, en estas versiones los índices de consistencia interna y los porcentajes de varianza explicada fueron los más altos. Se concluyó que el uso de seis y siete opciones de respuesta, específicamente en escalas psicométricas breves y unidimensionales, ofrece mediciones más precisas, en comparación con las de cinco. Se discuten las implicaciones de estos hallazgos.

**Palabras Clave:** *Número de opciones de respuesta, escalas Likert, psicometría, confiabilidad, validez.*

**Abstract:**

Psychosocial scales are used profusely in studies in many different fields. However, psychometric research has not been conclusive regarding the number of response options that is optimal in measuring a construct. The objective of the present study was to compare the psychometric properties of four psychosocial scales using different number of response options. The participants (757 university students) were assigned to one of three independent groups to answer a battery consisting of four short scales that evaluated gratitude, optimism, subjective well-being and loneliness, with five, six or seven response options. A positive association was found between number of options and psychometric quality of the scales. The greatest variability (dispersion) was observed in versions of six and seven options; also, in these versions the internal consistency indices and the explained variance percentages were the highest. It was concluded that the use of six and seven response options, specifically at short and one-dimensional psychometric scales, offers more precise measurements, compared to five. The implications of these findings are discussed.

**Keywords:** *Number of response options, Likert scales, psychometrics, reliability, validity.*

---

<sup>1</sup>Universidad Nacional Autónoma de México, México

\*Correspondencia: [ncalleja@unam.mx](mailto:ncalleja@unam.mx)

Los instrumentos psicométricos de auto reporte con escalas de respuesta tipo Likert constituyen herramientas de medición utilizadas profusamente en estudios de áreas tan diversas como la psicología, la mercadotecnia, la educación, la salud, la industria, la economía y la sociología. Una escala se conforma con varios reactivos integrados por una afirmación declarativa y tres o más opciones de respuesta, que permiten a los respondientes graduar sus respuestas con descriptores verbales (como “Totalmente de acuerdo”- “Totalmente en desacuerdo” o “Siempre”- “Nunca”) y/o con números; para efectuar los análisis, se promedian los valores cuantitativos o puntajes numéricos asignados a cada opción de respuesta (DeVellis, 2017; Maeda, 2015; Mills, & Gay, 2016). Es frecuente que los investigadores se cuestionen acerca de las propiedades psicométricas de las escalas que aplican y que indaguen cómo hacer que sus mediciones sean cada vez más precisas (Croasmun & Ostrom, 2011; Revilla, Saris, & Krosnick, 2014). Entre las numerosas decisiones que deben tomarse al aplicar escalas Likert se encuentra el número de opciones de respuesta. Aunque con frecuencia se utilizan cuatro o cinco (Preston & Colman, 2000), la literatura sobre construcción de instrumentos psicosociales ha mostrado que no necesariamente constituyen un número óptimo.

El número de opciones de respuesta tiene que ver con el nivel de medición de las variables. Desde que Stevens (1946, 1957) propuso las cuatro escalas de medición en psicología (nominal, ordinal, intervalar y de razón), se ha discutido el nivel al que corresponden los instrumentos tipo Likert. En sentido estricto, debido a que el tamaño de los intervalos entre las opciones de respuesta no es exactamente el mismo, no se trata de mediciones intervalares y, por tanto, nunca podría

usarse estadística paramétrica con tales datos; no obstante, los investigadores generalmente tratan a nivel intervalar la mayoría de los tests y escalas multi-reactivos cuyos puntajes se suman o promedian (Furr, 2018). Miles y Shevlin (2011) proponen que la pregunta que se plantee el investigador respecto de la variable medida no sea: “¿está en una escala intervalar?”, sino “¿está suficientemente cerca de una escala intervalar? (p. 62), y ello hace alusión al número de categorías u opciones de respuesta que se usen en el instrumento. Al emplear el número máximo de categorías que los participantes pueden discriminar, se estará más cerca de una escala de intervalo.

Respecto del número ideal de opciones de respuesta, los investigadores mantienen una de tres posturas generales: a) al aumentar el número de opciones, se incrementa la validez y la confiabilidad del instrumento (v.gr., Allen & Seaman, 2007; Lozano, García-Cueto, & Muñiz, 2008; Weng, 2004), b) menos opciones mejoran la medición (v.gr., Chang, 1994; Revilla et al., 2014), y c) el número de opciones no afecta la calidad psicométrica de la escala (v.gr., Dawes, 2008; Domínguez, 2013).

En relación con los estudios que señalan que existe una asociación positiva entre el número de opciones y las propiedades psicométricas de la escala, diversos autores han encontrado que tener un mayor número de ellas permite obtener más información del individuo (Chang, 1994; Comrey, 1988, Garner, 1960; Green & Rao, 1970), y que más opciones de respuesta llevan a mayor evidencia de validez y mejor capacidad discriminativa (Nunnally & Bernstein, 1994; Preston & Colman, 2000). Lozano et al. (2008) reportaron que, conforme aumenta el número de opciones de respuesta en las escalas aplicadas, aumenta el porcentaje de varianza explicada.

Asimismo, se ha señalado que usar más opciones incrementa la confiabilidad obtenida, aunque hasta cierto punto, después del cual la confiabilidad no aumenta a pesar de que se agreguen más opciones de respuesta. En este sentido, algunos autores han encontrado que el punto en el que se maximiza la confiabilidad corresponde a siete opciones de respuesta (Allen & Seaman, 2007; McKelvie, 1978; Ramsay, 1973), y otros reportan que la confiabilidad se estabiliza utilizando cinco (Cummins & Gullone, 2000) o seis opciones (Aiken, 1983; Boote, 1981; Weng, 2004). Streiner, Norman y Cairney (2015) han señalado que existe evidencia empírica de que las personas prefieren y son capaces de discriminar siete opciones. Simms, Zelazny, Williams y Bernstein (2019) analizaron de dos a once opciones y encontraron que la precisión psicométrica era baja en los subtests del inventario de personalidad que aplicaron con cinco o menos opciones, pero que después de seis se mantenía sin variación. En cuanto al análisis multivariado de los datos, Martínez y Rodríguez (2017) realizaron un estudio de simulación para evaluar el efecto del número de opciones de respuesta en el cálculo de diferentes coeficientes de correlación. Sus análisis mostraron que la estimación producto-momento, en contraste con la estimación tetracórica-politórica, infravalora la relación entre las variables cuando el número de opciones de respuesta es pequeño (dos, tres o cuatro); sin embargo, ambos coeficientes resultan muy similares cuando el número de opciones de respuesta es mayor (cinco o siete). En otro estudio de simulación, Maydeu-Olivares, Fairchild y Hall (2017) mostraron que, puesto que con pocas opciones de respuesta la desviación estándar se reduce y la curtosis aumenta, el poder ( $1-\beta$ ) de los índice de ajuste

(como  $\chi^2$ , RMSEA y SRMR) para detectar modelos incorrectos en la estructura factorial del instrumento resulta severamente afectado cuando se utilizan menos de cinco opciones.

En contraposición a esta postura, Revilla et al. (2014) realizaron un estudio en el que compararon escalas con cinco, siete y once opciones de respuesta y sus resultados mostraron que un mayor número de opciones hizo que los participantes efectuaran más interpretaciones, lo cual condujo a obtener menor validez. Respecto de la confiabilidad, Chang (1994) comparó los coeficientes de consistencia interna de instrumentos con cuatro y seis opciones de respuesta y encontró que el cuestionario de cuatro opciones mostró mayores índices de confiabilidad que el de seis.

También se ha señalado que la validez y la confiabilidad del instrumento son independientes del número de opciones de respuesta (Schuts & Rucker, 1975). López (2005), al comparar cinco formatos obtenidos al “colapsar” las cuatro de opciones de respuesta de un test de depresión a dos o tres categorías, reportó la misma confiabilidad y validez con un formato politómico que con uno dicotómico. En un estudio similar, Domínguez (2013) aplicó una escala de autoeficacia con cuatro opciones de respuesta, analizó diferentes combinaciones de éstas, también agrupadas en tres y dos opciones, y llegó a la conclusión de que sus índices psicométricos no fueron afectados por este factor. Dawes (2008), al evaluar el impacto de cinco, siete y diez opciones de respuesta en la media, la desviación estándar, el sesgo y la curtosis, reportó que no hubo diferencias entre los instrumentos y concluyó que la información obtenida es útil, independientemente del número de opciones de respuesta; sin embargo, no comparó sus índi-

ces psicométricos. Wakita, Ueshima y Noguchi (2012) aplicaron un cuestionario de personalidad con cuatro, cinco y siete opciones de respuesta y señalaron que la confiabilidad fue independiente del número de categorías utilizadas en el cuestionario.

En virtud de que es fundamental medir con precisión en la psicología (y en todas las ciencias) y de que aún no existe consenso respecto del efecto del número de opciones de respuesta y del medio de aplicación de un instrumento sobre la calidad de los datos obtenidos, el objetivo del presente estudio fue comparar diferentes versiones de escalas de constructos psicológicos con formato de respuesta tipo Likert. Se hipotetizó que en las versiones con mayor número de opciones de respuesta se obtendrían índices psicométricos más altos (variabilidad, confiabilidad y validez factorial). A fin de evaluar el posible efecto diferencial de estas variaciones según el constructo medido, se aplicaron cuatro escalas breves unidimensionales de variables psicosociales: optimismo, gratitud, bienestar psicológico y soledad. El estudio pretende aportar evidencia empírica que apoye la toma de decisiones de los investigadores respecto del número de opciones más adecuado para utilizar en escalas que permita incrementar la validez y confiabilidad de sus mediciones. Adicionalmente, esta investigación tuvo el propósito de aportar evidencia empírica sobre esta temática en población mexicana.

## Método

### Participantes

La muestra fue de tipo no aleatorio intencional; estuvo constituida por 757 estudiantes universitarios de licenciatura (85.2%) y posgrado (14.8%), de instituciones públicas (83.9%) y privadas (16.1%) de la Ciudad de

México y área metropolitana. El 72.0% de los participantes eran mujeres; sus edades oscilaron entre los 18 y los 30 años ( $M=21.82$  años,  $DE=3.24$ ); la mayoría dijeron ser solteros (93.7%) y no tener un trabajo remunerado (69.9%).

### Diseño

Se empleó un diseño cuasiexperimental de tres grupos no aleatorios. El 30.02% de los participantes respondió la versión de cinco opciones de respuesta, el 35.63% la de seis y el 34.35% la de siete.

### Instrumentos

Se conformó una batería integrada por un cuestionario demográfico (que registraba sexo, edad, estado civil, trabajo remunerado, nivel de estudios y tipo de universidad) y cuatro escalas psicosociales, cuyos reactivos se muestran en el Apéndice: Cuestionario de Optimismo, Cuestionario de Gratitud, Escala de Bienestar Subjetivo y Escala Multifactorial de Soledad. Estas escalas fueron seleccionadas por ser instrumentos cortos (entre 7 y 10 reactivos), unidimensionales y con índices psicométricos adecuados, características que permitían efectuar los análisis requeridos para el cumplimiento de los objetivos planteados y evaluar el efecto de las variables del estudio. Además, los constructos medidos resultaron de interés y fueron pertinentes para los integrantes de la muestra. El número total de reactivos para las cuatro escalas fue de 34.

*Cuestionario de Optimismo, COP* (Pedrosa, Celis, Suárez, García y Muñiz, 2015). Se dirige a jóvenes y está integrado por un solo factor, que explica el 42.43% de la varianza total, con nueve reactivos y cinco opciones de respuesta; los autores reportaron un índice de consistencia interna  $\alpha$  de Cronbach = .84. Mi-

de el optimismo disposicional, es decir, la tendencia personal estable de evaluar positivamente los posibles eventos futuros (Carver, Scheier y Segerstrom, 2010).

*Cuestionario de Gratitude, GQ-6* (McCullough, Emmons y Tsang, 2002). El instrumento original, Gratitude Questionnaire-6, se dirige a estudiantes universitarios y consta de un solo factor, conformado por seis reactivos con siete opciones de respuesta; los autores probaron el ajuste unidimensional de los datos mediante un análisis factorial confirmatorio (CFI=0.95, SRMR=0.04) y reportaron un coeficiente alfa de .82. El GQ-6 evalúa la gratitud como una disposición para reconocer las acciones de benevolencia de otras personas y responder a ellas con una emoción de agradecimiento. El GQ-6 fue sometido al proceso de traducción, re-traducción y adaptación por los autores del presente trabajo. La versión en español quedó integrada finalmente por siete reactivos, ya que el reactivo 5. “As I get older I find myself more able to appreciate the people, events, and situations that have been part of my life history”, fue dividido en dos: “Conforme avanzo en edad, valoro más a las personas que han formado parte de mi vida” y “Conforme soy mayor, valoro más las experiencias que he vivido”.

*Escala de Bienestar Subjetivo (EBS)* (Calleja y Almazán, en prensa). Fue desarrollada con el propósito de evaluar el bienestar subjetivo en jóvenes mexicanos. En su construcción se observaron los procedimientos psicométricos establecidos (cfr. DeVellis, 2017). El constructo se definió como la evaluación que hacen las personas de su satisfacción con la vida (con su trabajo, su escuela, su familia), cognoscitiva y emocionalmente (Diener, 2006; Diener y Diener, 1996) y se diferenció del bienestar eudomonista y del bienestar psico-

lógico (González y Andrade, 2016). Consta de una dimensión con ocho reactivos y siete opciones de respuesta de acuerdo-desacuerdo. Su validación en población general mostró una consistencia interna  $\alpha=.971$  y una varianza explicada de 78.73%. Los índices de ajuste del análisis factorial confirmatorio resultaron satisfactorios (CMIN/DF = .668; GFI = .980, NFI = .992; CFI = .998; RMSEA =.041).

*Escala Multifactorial de Soledad* (Montero, 1998). Está conformada por cuatro factores, que explican el 51% de la varianza total; los índices de confiabilidad de los factores oscilan entre .72 y .94. En el presente estudio se aplicaron únicamente los 10 reactivos con mayor carga factorial de la subescala de Bienestar Emocional. La soledad se ha definido como un fenómeno psicológico potencialmente estresante, que resulta de percibir carencias afectivas, sociales y físicas (Montero y Sánchez, 2001).

### Procedimiento

Con el objetivo de evitar sesgos en las respuestas, los reactivos de las cuatro escalas fueron integrados y ordenados de manera aleatoria en una batería. Se crearon tres versiones de la batería, una en la que los reactivos tenían cinco opciones de respuestas, otra con seis y la tercera con siete. En esta última, las opciones fueron: Muy en desacuerdo (0), En desacuerdo (1), Ni en acuerdo ni en desacuerdo (2), De acuerdo (3), Muy de acuerdo (4), Bastante de acuerdo (5) y Totalmente de acuerdo (6); en la versión de seis opciones se eliminó la de Muy en desacuerdo, y en la de cinco, además de ésta, la de Bastante de acuerdo (véase tabla 1). La batería se elaboró tanto en formato impreso como en versión electrónica. Cada uno fue piloteado con 15 voluntarios que poseían características

similares a los de la muestra. Inmediatamente después de la aplicación, se efectuaron entrevistas estructuradas para explorar dificultades en la comprensión de cada uno de los reactivos, errores en la secuencia de los mismos y problemas con el formato de respuesta. Los datos obtenidos fueron analizados por el grupo de investigadores responsables; se decidió realizar los siguientes cambios: invertir y modificar el fraseo del reactivo 3 de la Escala de Optimismo (“Pienso que todo saldrá mal” por “Tiendo a pensar que todo saldrá bien”), del reactivo 3 de la Escala de Gratitud (“Cuando miro al mundo, no veo mucho por lo que estar agradecido” por “Cuando miro a mi alrededor, considero que hay mucho por lo que estar agradecido”) y del 6 de la misma escala (“Puede pasar mucho tiempo antes de que me sienta agradecido por algo o por alguien” por “Con frecuencia me siento agradecido(a) por algo o con alguien”). Asimismo, se cerraron todas las preguntas del cuestionario sociodemográfico. No fue necesario modificar el formato ni la secuencia de los reactivos.

La batería se aplicó tanto en formato impreso (59.45%) como electrónico (40.55%). Los instrumentos impresos fueron aplicados a los estudiantes universitarios en sus propios salones de clase o en el campus de la institución a la que asistían; en el primer caso, en forma grupal y en el segundo, de manera individual. Después de presentarse, los aplicadores explicaban los objetivos del estudio y los invitaba a participar en él de manera voluntaria; se hacía énfasis en que sus respuestas serían tratadas de forma completamente anónima y confidencial. Las aplicaciones electrónicas fueron respondidas en las plataformas Google Forms y E-survey creator, y en los formularios de Adobe Acrobat IX; la difusión se efectuó a través de redes sociales o por invitación directa vía correo electrónico.

### **Análisis de los datos**

Las propiedades psicométricas de los cuatro instrumentos que conformaron la batería fueron analizadas separadamente por versión de opciones de respuesta (cinco, seis y siete). Se efectuaron los siguientes análisis estadísticos: distribución de frecuencias en las diferentes opciones de respuesta, medias y desviaciones estándar, pruebas t para muestras independientes y análisis de varianza (Anova de un factor) y análisis de confiabilidad mediante índices de consistencia interna ( $\alpha$  de Cronbach). Para confirmar la estructura factorial de las escalas, evaluar los índices de ajuste así como determinar el porcentaje de varianza media extraída (AVE), se efectuaron análisis factoriales confirmatorios (AFC) con el método de máxima verosimilitud. Se analizaron los siguientes índices (Hu & Bentler, 1999): prueba de bondad de ajuste  $\chi^2$ ; cociente  $\chi^2/\text{gl}$  (CMIN/DF) cuyo valor  $<3$  indica un ajuste adecuado, los índices comparativos CFI (Comparative Fit Index) y TLI (Tucker Lewis Index), que con valores  $\geq .95$  se interpretan como propios de un ajuste adecuado; así como los de error: SRMR (Standardized Root Mean Square Residual) y RMSEA (Root Mean Square of Aproximation), que muestran un ajuste adecuado con valores  $< .08$  en el primer caso, y  $< .06$ , en el segundo. Respecto de la varianza media extraída, Fornell y Larcker (1981) recomiendan valores por encima de .50. Se emplearon los programas SPSS y AMOS, ambos versión 22, para llevar a cabo los procedimientos estadísticos.

### **Resultados**

#### *Distribución de frecuencias*

Al analizar la distribución de frecuencias en cada una de las versiones de opciones de respuesta y para cada uno de los instrumentos aplicados, se observó una mayor dispersión

en la versión de siete opciones de respuesta y la mayor concentración en una de las opciones de respuesta en la de cinco; sin embargo, en ningún caso rebasó el 50%. En la tabla 1 se presenta el porcentaje promedio de participantes en cada una de las opciones de respuesta de los reactivos que integraron cada uno de los instrumentos, en las versiones de cinco, seis y siete opciones. Para cada instrumento, el promedio por opción se obtuvo sumando a los participantes que eligieron esa

opción en todos los reactivos de la escala, y dividiendo el total entre el número de reactivos. En todos los casos, el porcentaje más alto en una opción de respuesta se encontró en la versión de cinco opciones, seguida por la de seis y, finalmente, la de siete. En los instrumentos de atributos positivos, los mayores porcentajes se ubicaron en las opciones de Muy de acuerdo a Totalmente de acuerdo o de Casi siempre y Siempre, y en la de Soledad, en la de En desacuerdo.

Tabla 1.

*Porcentaje promedio de participantes en cada una de las opciones de respuesta de las escalas y subescalas aplicadas.*

Versión por número de opciones	Opciones de respuesta							Total
	Muy en des acuerdo	En desacuerdo	Ni en acuerdo ni en desacuerdo	De acuerdo	Muy de acuerdo	Bastante de acuerdo	Totalmente de acuerdo	
	0	1	2	3	4	5	6	
<i>Cuestionario de Optimismo (9 reactivos)</i>								
5	—	1.24	10.04	28.93	<b>32.71</b>	—	27.08	100
6	—	2.41	10.37	19.44	17.00	23.96	<b>26.81</b>	100
7	1.84	3.97	10.64	20.60	16.97	<b>24.91</b>	21.07	100
<i>Cuestionario de Gratitud (7 reactivos)</i>								
5	—	1.31	7.40	23.21	30.53	—	<b>37.56</b>	100
6	—	1.05	4.76	15.43	15.95	23.00	<b>39.81</b>	100
7	1.32	3.24	6.54	17.69	15.55	23.46	<b>32.20</b>	100
<i>Escala de Bienestar Subjetivo (8 reactivos)</i>								
5	—	0.51	9.05	24.37	—	<b>39.00</b>	27.07	100
6	—	1.26	6.49	16.97	17.47	28.87	<b>28.94</b>	100
7	1.1	3.4	9.2	18.5	16.8	<b>27.7</b>	23.4	100
<i>Escala de Soledad (10 reactivos)</i>								
5	—	<b>49.85</b>	41.83	5.28	—	2.18	0.86	100
6	—	<b>35.24</b>	33.06	22.67	5.17	3.13	0.73	100
7	31.85	<b>33.69</b>	22.19	5.46	3.00	2.27	1.54	100

### *Medias y desviaciones estándar*

Se probó la significancia de las diferencias entre los grupos que respondieron a diferentes opciones de respuesta. Las desviaciones estándar más altas (mayores dispersiones) se encontraron en la versión de siete opciones y las medias más altas en la de seis. Las diferencias resultaron significativas en dos de las cuatro escalas (Optimismo y Gratitud); en ellas, la versión de cinco difirió significativamente de la de seis opciones y la de seis de la de siete (véase tabla 2).

### *Índices de consistencia interna*

Los índices más altos de consistencia interna de los instrumentos aplicados correspondieron a la versión de siete opciones, con excepción de la Escala de Bienestar Subjetivo, en la que el alfa de Cronbach más alto se ubicó en la versión de seis opciones. En la escala de Optimismo y en la de Gratitud la mayor diferencia ocurrió entre la versión de cinco y la de siete opciones de respuesta, y la menor, entre la versión de seis y siete opciones (véase tabla 3). En tres de los cuatro instrumentos, las menores diferencias en los índices alfa se presentaron entre la versión de seis y la de siete opciones.

### *Estructura factorial*

#### *y varianza media extraída*

Mediante análisis factoriales confirmatorios se probaron las soluciones unifactoriales para las cuatro escalas. La varianza media extraída (que indica la proporción de la varianza de los reactivos explicada por el factor latente) resultó  $>.50$  en las versiones de seis y de siete opciones, siendo más alta en éstas últimas, excepto para la Escala de Bienestar Subjetivo. Si bien en todos los casos el cociente  $\chi^2/df$  fue  $<.3$ , los menores cocientes se obtuvieron en las versiones de siete opciones. Los índices de ajuste comparativo (TLI y CFI) superaron

el criterio de .95 en las versiones de seis y siete opciones y fueron más altos en los de siete, en tanto que los menores índices de error (SRMR y RMSEA) se obtuvieron con siete opciones, seguidas por las de seis. Las diferencias más altas ocurrieron entre las versiones de cinco y siete opciones, y las menores entre las de seis y siete.

### **Discusión**

La finalidad del presente estudio fue determinar el efecto del número de opciones de respuesta (cinco, seis o siete) de escalas tipo Likert sobre la calidad psicométrica y los resultados de la medición de distintos constructos. Los criterios de evaluación fueron: variabilidad, índice de consistencia interna (confiabilidad), porcentaje de la varianza explicada (validez factorial) y estadísticos descriptivos.

En todas las escalas se encontró la mayor dispersión de participantes en las versiones de siete opciones y, consecuentemente, la mayor concentración en una de las opciones de respuesta en las versiones de cinco opciones. Sin embargo, los porcentajes no rebasaron el 50%, lo que probablemente ocurriría en versiones de tres o cuatro opciones. Los conceptos fundamentales en medición psicológica se construyen sobre el concepto de variabilidad (Furr, 2018), ya que ésta refleja las diferencias en la cantidad del atributo que poseen las personas evaluadas. La desviación estándar, el indicador estadístico de la variabilidad, resultó mayor en la versión de siete opciones y en la aplicación electrónica versus la de papel, lo que muestra mayor precisión en la medición y, por tanto, mayor calidad psicométrica.

Se esperaba que las medias obtenidas fueran similares, independientemente del número de opciones de respuesta y del medio de aplicación que se utilizara. Esto ocurrió en dos de las cuatro escalas aplicadas (Bienestar Subjetivo y Soledad), en las que no se encontraron

Tabla 2.

Media\* (desviación estándar) para los grupos de las tres versiones de opciones de respuesta y prueba estadística de las diferencias entre ellos.

Instrumentos	Versión de número de opciones			Anova	Pruebas post hoc Scheffe
	5	6	7		
Optimismo	4.01 (0.96)	<b>4.30</b> (1.14)	4.06 ( <b>1.24</b> )	F(2,754) = 4.938, p = .007	5-6: p = .022 5-7: n.s. 6-7: p = .042
Gratitud	4.33 (0.98)	<b>4.75</b> (1.03)	4.42 ( <b>1.20</b> )	F(2,754) = 12.349, p < .001	5-6: p < .001 5-7: n.s. 6-7: p = .002
Bienestar subjetivo	4.24 (1.06)	<b>4.44</b> (1.22)	4.21 ( <b>1.25</b> )	F(2,754) = 2.940, n.s.	n.s.
Soledad	1.15 (0.91)	<b>1.12</b> (0.88)	1.27 ( <b>0.99</b> )	F(2,754) = 1.752, n.s.	n.s.

\*Media teórica = 3; rango 1 a 7; n.s.: Diferencia no significativa.

Tabla 3.

Índices de consistencia interna ( $\alpha$ ) de las escalas en las tres versiones de número de opciones y diferencias entre ellos.

Instrumentos	Versión de número de opciones			Diferencias
	Cinco	Seis	Siete	
Optimismo	.883	.932	<b>.937</b>	5-6: -0.049 <b>5-7: -0.054</b> 6-7: -0.005
Gratitud	.850	.908	<b>.910</b>	5-6: -0.058 <b>5-7: -0.060</b> 6-7: -0.002
Bienestar subjetivo	.922	<b>.952</b>	.944	<b>5-6: -0.030</b> 5-7: -0.022 6-7: 0.008
Soledad	.896	.905	<b>.925</b>	5-6: -0.009 <b>5-7: -0.029</b> 6-7: -0.020

Tabla 4.

Varianza Media Extraída (AVE) e índices de ajuste obtenidos en los análisis factoriales confirmatorios de las escalas en las tres versiones de número de opciones.

Instrumentos	AVE e índices de ajuste	Versión de número de opciones			Diferencias		
		Cinco	Seis	Siete	5-6	5-7	6-7
Optimismo	AVE	.41	.58	<b>.61</b>	-.17	<b>-.20</b>	-.03
	$\chi^2$ (gl=21)	49.817	46.341	<b>40.083</b>	3.476	<b>9.734</b>	6.258
	$\chi^2$ /gl	2.372	2.207	<b>1.908</b>	0.165	<b>0.464</b>	0.299
	TLI	.942	.974	<b>.983</b>	-.032	<b>-.041</b>	-.009
	CFI	.966	.985	<b>.989</b>	-.019	<b>-.023</b>	-.004
	SRMR	.041	.026	<b>.024</b>	.015	<b>.017</b>	.002
	RMSEA	.078	.068	<b>.056</b>	.010	<b>.022</b>	.012
Gratitud	AVE	.44	.58	<b>.59</b>	-.14	<b>-.15</b>	-.01
	$\chi^2$ (gl=12)	32.748	25.277	<b>19.716</b>	7.471	<b>13.032</b>	5.561
	$\chi^2$ /gl	2.949	2.106	<b>1.643</b>	0.843	<b>1.306</b>	0.463
	TLI	.895	.981	<b>.988</b>	-.086	<b>-.093</b>	-.007
	CFI	.955	.989	<b>.994</b>	-.034	<b>-.039</b>	-.005
	SRMR	.044	.022	<b>.021</b>	.022	<b>.023</b>	.001
	RMSEA	.094	.061	<b>.050</b>	.033	<b>.044</b>	.011
Bienestar subjetivo	AVE	.55	<b>.67</b>	.62	-.12	<b>-.07</b>	.05
	$\chi^2$ (gl=14)	39.091	30.236	<b>24.722</b>	8.855	<b>14.369</b>	5.514
	$\chi^2$ /gl	2.792	2.160	<b>1.766</b>	0.632	<b>1.026</b>	0.394
	CFI	.977	.984	<b>.994</b>	-.007	<b>-.017</b>	-.010
	TLI	.967	.978	<b>.988</b>	-.011	<b>-.021</b>	-.010
	SRMR	.036	.026	<b>.019</b>	.010	<b>.017</b>	.007
	RMSEA	.077	.067	<b>.054</b>	.010	<b>.023</b>	.013
Soledad	AVE	.47	.52	<b>.56</b>	-.05	<b>-.09</b>	-.04
	$\chi^2$ (gl=31))	80.536	58.977	<b>53.219</b>	21.559	<b>27.317</b>	5.758
	$\chi^2$ /gl	2.598	1.902	<b>1.717</b>	0.696	<b>0.881</b>	0.185
	TLI	.915	.975	<b>.979</b>	-.060	<b>-.064</b>	-.004
	CFI	.942	.983	<b>.986</b>	-.041	<b>-.044</b>	-.003
	SRMR	.046	.030	<b>.028</b>	.016	<b>.018</b>	.002
	RMSEA	.090	.055	<b>.053</b>	.035	<b>.037</b>	.002

Índices de ajuste y criterios de corte:  $\chi^2$ /gl <3; TLI (Tucker Lewis Index)  $\geq$ .95; CFI (Comparative Fit Index)  $\geq$ .95; SRMR (Standarized Root Mean Square Residual) <.08; RMSEA (Root Mean Square of Aproximation) <.06.

diferencias significativas entre las medias de las versiones de cinco, seis y siete opciones. Wakita et al. (2012) no encontró diferencias en su medición con cuatro y cinco opciones de respuesta, pero sí con siete. Leung (2011) reportó resultados similares entre las medias de instrumentos con 4, 5, 6 y 11 opciones de respuesta. Habrá que evaluar, en un estudio posterior, la razón por la que en el presente estudio se obtuvieron diferencias en dos escalas (Optimismo y Gratitud) y por qué la media de la versión de seis opciones resultó la más alta en todos los casos. Entre tanto, habrá que tener muy presente este efecto diferencial del número de opciones sobre las medias.

Los índices de consistencia interna más altos de tres de los cuatro instrumentos aplicados correspondieron a las versiones de siete opciones (en la escala de Bienestar Subjetivo el índice alfa más alto se obtuvo en seis opciones, aunque la diferencia con la de siete fue mínima). Estos resultados concuerdan con los reportados por otros autores (v.gr., Cicchetti et al., 1985; McKelvie, 1978; Preston & Colman, 2000), quienes afirman que la confiabilidad tiende a maximizarse con el uso de seis y siete opciones de respuesta.

En cuanto a la validez factorial, el mayor porcentaje de varianza media extraída fue obtenido en las versiones de siete opciones de respuesta y el menor, en las de cinco (excepto, nuevamente, en la Escala de Bienestar Subjetivo, en la que el índice más alto correspondió a la versión de seis opciones, aunque la diferencia con la de siete fue mínima). Lozano et al. (2008) también encontraron que al aumentar el número de opciones de respuesta, la varianza explicada se incrementa. En relación con el ajuste de los modelos factoriales a los datos, en los análisis factoriales confirmatorios se obtuvieron consistentemente mejores índices en las versiones de siete y seis opciones que en las de cinco, para las

cuatro escalas. Es decir, con instrumentos que sólo difieren en el número de opciones de respuesta, los procedimientos estadísticos arrojan mejores ajustes para las mediciones más finas que para las más gruesas, debido a la mayor variabilidad que es posible obtener. Puesto que, como señalan Maydeu-Olivares et al. (2017), el modelamiento estructural con variables latentes es una piedra angular de la psicometría y se usa con frecuencia para probar y construir teorías, es fundamental tomar en cuenta que el poder que tienen para probar los modelos propuestos está influenciado de manera decisiva por el número de opciones de respuesta.

### Limitaciones y propuestas

Los resultados de este estudio indican que las escalas con seis y siete opciones de respuesta presentan, en términos generales, índices psicométricos más fuertes que las de cinco. Además, al utilizar este número de opciones de respuesta se obtiene mayor variabilidad, se realizan distinciones más precisas en la medición del atributo y es factible aplicar procedimientos estadísticos paramétricos.

Sin embargo, en virtud de que la investigación fue realizada en población universitaria, que suele estar familiarizada con cuestionarios y encuestas y que posee habilidades para discriminar entre varias opciones de respuesta, será necesario realizar estudios semejantes con muestras de características heterogéneas. Otra limitación es que sólo se analizaron cinco, seis y siete opciones de respuesta; sería conveniente incluir menos (dos, tres y cuatro) y más opciones (ocho, nueve y diez) para lograr una comprensión más completa del fenómeno. Asimismo, puesto que en el presente estudio se evaluaron tres emociones positivas, las cuales tienen efectos específicos

sobre el comportamiento (Schmidt, 2008) y cuyos puntajes se ubicaron por arriba de la media teórica, es probable que difieran de otras variables psicológicas, por lo que se sugiere replicarlo midiendo otros constructos y/o utilizando otros instrumentos. También habrá que considerar que las escalas aplicadas fueron cortas y unidimensionales (de siete a diez reactivos), por lo que es posible que los resultados varíen con instrumentos de mayor extensión. Asimismo, se requiere evaluar la comprensión que los participantes tienen de las diferentes opciones de respuesta y preguntarles su preferencia respecto de éstas. Finalmente, habrá que considerar que los elementos psicométricos estudiados aquí son solo un aspecto del extraordinariamente complejo proceso de la medición del comportamiento humano.

### Referencia

- Aiken, L. R. (1983). Number of response categories and statistics on a teacher rating scale. *Educational and Psychological Measurement*, 43(2), 397-401. doi: 10.1177/001316448304300209
- Allen, E., & Seaman, C. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64-65. Recuperado de: <http://rube.asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html>
- Boote, A. S. (1981). Reliability testing of psychographic scales. *Journal of Advertising Research*, 21(5), 53-60. doi: <http://psycnet.apa.org/record/1982-04609-001>
- Calleja, N. y Almazán, T. (en prensa). Escala de Bienestar Subjetivo (EBS-20 y EBS-8): Construcción, validación e invarianza de medición. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*.
- Carver, C. S., Scheier, M. F., & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review*, 30(7), 879-889. doi: 10.1016/j.cpr.2010.01.006
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), 205-215. doi: 10.1177/014662169401800302
- Cicchetti, D.V., Showalter, D., & Tyrer, P.J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement*, 9(1), 31-36. doi: 10.1177/014662168500900103
- Comrey, A. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56(5), 754-761. doi:10.1037/0022-006X.56.5.754
- Croasmun, J. & Ostrom, L. (2011). Using Likert-type scales in the Social Sciences. *Journal of Adult Education*, 40(1), 19-22. <https://files.eric.ed.gov/fulltext/EJ961998.pdf>
- Cummins, R. A. & Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality-of-life measurement. In *Proceedings of the Second International Conference on Quality of Life in Cities* (pp. 74-93). Kent Ridge, Singapore: National University of Singapore.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1), 61-104. doi: 10.1177/147078530805000106
- DeVellis, R. F. (2017). *Scale development: Theory and applications*. (4th ed.). Los Angeles, CA: Sage.
- Diener, E. & Diener, C. (1996). Most people are happy. *Psychological Science*, 7(3), 181-185. doi: 10.1111/j.1467-9280.1996.tb00354.x.
- Diener, E. (2006). Guidelines for national indicators of subjective wellbeing and ill being. *Journal of Happiness Studies*, 7(4), 397-404. doi: 10.1007/s10902-006-9000-y
- Domínguez, S. A. (2013). ¿Ítems politómicos o dicotómicos? Un estudio empírico con una escala unidimensional. *Revista Argentina de Ciencias del Comportamiento*, 5(3), 30-37. Recuperado de: <https://revistas.unc.edu.ar/index.php/racc/article/view/5562/Lara>.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. doi: 10.2307/3151312
- Furr, R. M. (2018). *Psychometrics: An introduction*. (3rd Edition). Los Angeles, CA: Sage.
- Garner, W. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, 67(6), 343-352. doi: 1037/h0043047
- González, M., & Andrade, P. (2016). Escala de Bienestar Psicológico para Adolescentes. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 42(2), 69-83. doi: 10.21865/ridep42\_69
- Green, P. & Rao, V. (1970). Rating scale and information recovery: How many scales and response category to use? *Journal of Marketing*, 34(3), 33-39. doi: 10.2307/1249817
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi: 10.1080/10705519909540118
- Maeda, H. (2015). Response option configuration of online administered Likert scales. *International Journal of*

- Social Research Methodology*, 18(1), 15-26. doi: 10.1080/13645579.2014.885159
- Martínez, F. & Rodríguez, M. J. (2017). Comportamiento de las correlaciones producto-momento y tetracórica-policórica en escalas ordinales: Un estudio de simulación. *Relieve*, 23(2), 11-19. doi: 10.7203/relieve.23.2.9476
- Leung, S. (2011). A Comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, 37(4), 412-421. doi: 10.1080/01488376.2011.580697
- López, J. (2005). Items politómicos vs. dicotómicos: un estudio metodológico. *Anales de Psicología*, 21(2), 339-344. Recuperado de: <http://www.redalyc.org/html/167/16721216/>
- Lozano, L., García-Cueto, E. & Muñoz, J. (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
- Maydeu-Olivares, A., Fairchild, A., & Hall, A. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling*, 24(4), 495-505. doi:10.1080/10705511.2017.1289816
- Mills, G. E., & Gay, L. R. (2016). *Educational research: Competencies for analysis and applications*. Upper Saddle Ridge, NJ: Pearson.
- McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, 82(1), 112-127. doi:10.1037//0022-3514.82.1.112
- McKelvie, S. J. (1978). Graphic rating scales. How many categories? *British Journal of Psychology*, 69, 185-202. doi:10.1111/j.2044-8295.1978.tb01647.x
- Miles, J. & Shevlin, M. (2011). *Applying regression & correlation: A guide for students and researchers*. Los Angeles, CA: Sage.
- Montero, M. E. y Sánchez, J. J. (2001). La soledad como fenómeno psicológico: Un análisis conceptual. *Salud Mental*, 24(1), 19-27. Recuperado de: <http://www.medigraphic.com/pdfs/salmen/sam-2001/sam011d.pdf>
- Montero, M. E. (1998). *Soledad: Desarrollo y validación de un inventario multifacético para su medición*. (Tesis de Doctorado). Recuperado de: <http://132.248.9.195/pdbis/269524/Index.html>
- Nunnally, J. & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Pedrosa, I., Celis-Atenas, K., Suárez-Álvarez, J., García-Cueto, E., & Muñoz, J. (2015). Cuestionario para la evaluación del optimismo: Fiabilidad y evidencias de validez. *Terapia psicológica*, 33(2), 127-138. doi: 10.4067/S0718-48082015000200007
- Preston, C. & Colman, A. (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
- Ramsay, J. O. (1973). Effects of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, 38, 513-532. doi: 10.1007/BF02291492
- Revilla, M., Saris, W., & Krosnick, J. (2014). Choosing the number of categories in Agree-Disagree Scales. *Sociological Methods & Research*, 43(1), 73-97. doi: 10.1177/0049124113509605
- Schmidt, C. (2008). Construcción de un cuestionario de emociones positivas en población entreterriana. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 26(2), 117-139. Recuperado de: <http://www.redalyc.org/html/4596/459647347007/>
- Schuts, H. & Rucker, M. (1975). A comparison of variables configuration across scale lengths: an empirical study. *Educational and Psychological Measurement*, 35(2), 319-324. doi: 10.1177/001316447503500210
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4), 557-566. doi: 10.1037/pas0000648
- Stevens, S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680. Recuperado de: [http://gaius.fpce.uc.pt/niips/novoplano/mip1/mip1\\_201314/scales/Stevens\\_1946.pdf](http://gaius.fpce.uc.pt/niips/novoplano/mip1/mip1_201314/scales/Stevens_1946.pdf)
- Stevens, S. (1957). On the psychophysical law. *Psychological Review*, 64(3), 153-181. doi: 10.1037/h0046162
- Streiner, D., Norman, G., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use*. (5ª ed.). Oxford: Oxford University Press.
- Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale. *Educational and Psychological Measurement*, 72(4), 533-546. doi: 10.1177/0013164411431162
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient  $\alpha$  test-retest reliability. *Educational and Psychological Measurement*, 64, 956-972. doi: 10.1177/0013164404268674

**Apéndice: Reactivos de las escalas aplicadas**

<i>Cuestionario de Optimismo – COP</i> (Pedrosa et al., 2015)	<i>Cuestionario de Gratitud</i> (McCullough et al., 2002)
<p>Tengo confianza en mi futuro.                      Tengo confianza en superar los problemas.                      Tiendo a pensar que todo saldrá bien.                      Veo los aspectos positivos de las cosas.                      Creo que conseguiré las principales metas de mi vida.                      Veo cada reto como una oportunidad de éxito.                      Me ocurren más cosas buenas que malas.                      Por muy mal que salgan las cosas encuentro aspectos positivos.                      Cuando pienso en el futuro soy positivo.</p>	<p>Conforme avanzo en edad, valoro más a las personas que han formado parte de mi vida.                      Tengo mucho que agradecer en la vida.                      Cuando miro a mi alrededor, considero que hay mucho por lo que estar agradecido.                      Conforme soy mayor, valoro más las experiencias que he vivido.                      Estoy agradecido(a) con una gran diversidad de personas.                      Con frecuencia me siento agradecido(a) por algo o con alguien                      Si tuviera que hacer una lista de todo lo que agradezco, sería una lista muy larga.</p>
<i>Escala de Bienestar Subjetivo (EBS)</i> (Calleja y Almazán, en prensa)	<i>Escala de Soledad</i> (Montero, 1998)
<p>Mi vida es feliz.                      Estoy satisfecho(a) con mi vida.                      Mi vida es maravillosa.                      Me gusta mi vida.                      Disfruto de mi vida.                      Soy una persona feliz.                      Mi vida me trae alegría.                      Estoy "de buenas".</p>	<p>Me siento desamparado(a).                      La soledad es mi única compañera.                      Me siento rechazado(a).                      Me siento solo(a).                      Me siento incomprendido(a).                      Me siento apartado(a).                      Me siento aislado(a).                      Me siento encerrado(a) en mí mismo(a).                      Siento que solo cuento conmigo mismo(a).                      Me siento abandonado(a).</p>