

Artículos

Invarianza en escala Likert de cinco y siete puntos del SWLS en 5 países iberoamericanos

Measurement Invariance in Five-Point and Seven-Point Likert Scale of the SWLS in 5 Ibero-American Countries

Marco Alberto Núñez Ramírez¹, Karla Alejandra Garduño Realivazquez² e Irma Guadalupe Esparza García¹

1 Instituto Tecnológico de Sonora

2 Universidad de Sonora

Resumen

La escala Likert se ha convertido en un método popular de evaluación dentro de las ciencias sociales, sin embargo, no existe un consenso sobre cuál opción de respuesta es mejor (de cinco o siete puntos). La presente investigación trató de comparar las propiedades psicométricas de ambas opciones de respuesta Likert mediante el caso de un instrumento de satisfacción con la vida. A través del análisis de invarianza de medida, se llevó a cabo un estudio cuantitativo y transversal. Con dos muestras no probabilísticas de jóvenes de cinco países hispanohablantes ($n_1 = 706$, $n_2 = 911$), se realizó un análisis factorial confirmatorio multigrupo (país y género), considerando la invarianza configural, métrica, escalar y estricta. Los hallazgos sugieren, con base en estudios previos, que la escala Likert-5 obtuvo resultados más favorables. Por lo tanto, el instrumento sí fue invariante al comparar países; no obstante, respecto al género, se obtuvieron algunos indicadores favorables para soportar la invarianza de medida del instrumento.

Palabras clave: satisfacción con la vida, escala Likert, invarianza de media, análisis factorial confirmatorio, validez

Autores

Marco Alberto Núñez Ramírez. Instituto Tecnológico de Sonora.

ORCID: <https://orcid.org/0000-0001-5825-4482>

Karla Alejandra Garduño Realivazquez. Universidad de Sonora.

ORCID: <https://orcid.org/0000-0002-5199-9163>

Irma Guadalupe Esparza García. Instituto Tecnológico de Sonora.

ORCID: <https://orcid.org/0000-0002-3715-0761>

Autor para correspondencia: Karla Alejandra Garduño Realivazquez e-mail: karla.realivazquez@gmail.com

Abstract

The Likert scale has become a popular method of assessment within the social sciences; however, there is no consensus on which response option is better (five-point or seven-point). The present research attempted to compare the psychometric properties of the two Likert response options, using a life satisfaction instrument. A quantitative and cross-sectional study was carried out by analyzing measurement invariance. Based on two non-probabilistic samples of young people from five Spanish-speaking countries (n1= 706, n2= 911), a multigroup (country and gender) confirmatory factor analysis was performed, considering configural, metric, scalar and strict invariance. Based on previous studies and the results obtained, the findings suggest that the Likert-5 scale obtained more favorable results. Therefore, the instrument was invariant when comparing countries; nonetheless, with respect to gender, some favorable indicators were obtained to support the measurement invariance of the instrument.

Key words: life satisfaction, Likert scale, measurement invariance, confirmatory factor analysis, validity

DOI <https://doi.org/10.36793/psicumex.v14i1.605>

Recibido 24 – Noviembre – 2022

Aceptado 21 – Febrero – 2024

Publicado 10 – Junio – 2024



Introducción

La escala Likert se ha convertido en un método de evaluación de uso popular dentro de la comunidad académica (Hernández et al., 2010), en especial dentro de las ciencias sociales (Croasmun y Ostrom, 2011; Simms et al., 2019). Esta fue creada por Rensis Likert (1932), y se basa en la calificación de afirmaciones (Asociación Americana de Psicología [APA], 2010), donde el participante, con el fin de medir actitudes, responde a cuestiones con valores que van desde una percepción en desacuerdo a otra que puede ser de acuerdo (Kerlinger y Lee, 2002); es decir, de un nivel desfavorable a lo favorable (Garwood, 2006). A través de esta escala es posible estudiar la percepción acerca de eventos, objetos y personas (Taherdoost, 2019).

Por otro lado, es importante resaltar que la escala Likert, como todo método de medición, posee ciertas ventajas y desventajas (Hartley, 2014). Una de sus ventajas es que ha demostrado ser atractiva para los participantes de las investigaciones (Garwood, 2006), además de que su administración es fácil (Preston y Colman, 2000) y su codificación es sencilla al momento de acumular un conjunto de datos (Heo et al., 2022); permitiendo recopilar y analizar ítems basados en rasgos latentes (Anjaria, 2022). Y, al ser un dato numérico, es posible generar inferencias estadísticas, las cuales suelen mostrar niveles adecuados de fiabilidad (Li, 2013).

Singh et al. (2021) enlista tres desventajas de la escala Likert que son: (1) solo mide la dirección de la actitud y no la intensidad de esta, (2) las puntuaciones utilizadas de dos participantes con actitudes opuestas/diferentes pueden llegar a sumar igual y (3) el análisis del ítem se basa en la media o mediana de las personas participantes. Argumentos coincidentes con los de Li (2013), quien afirma que pueden existir problemas en relación con la pérdida y distorsión de la información; esto debido a su naturaleza ordinal y a que posee un formato cerrado. Además, Anjaria (2022) señala que al contar con tantas opciones de respuesta



puede ser imposible para los participantes describir adecuadamente su opinión con relación a su sentimiento real.

Debido a los inconvenientes de la escala Likert, durante varias décadas se ha originado un debate sobre el uso apropiado de la escala Likert (Carifio y Perla, 2008). Aunque el número de opciones de esta escala varía de 2 a 100 (Cummins y Gullone, 2000), y algunas escalas puedan ser adaptadas a un mayor número de respuestas con el fin de mejorar la distinción de las posibles respuestas de los participantes (Bryman, 2003), también han surgido diferentes controversias. Por ejemplo, existen discusiones sobre cuál escala es mejor (Colman et al., 1997; Cummins y Gullone, 2000). De hecho, Cox III (1980), luego de hacer una revisión entre los años 1900 y 1980, encontró que no existe una respuesta apropiada para todas las circunstancias. Desde los estudios de Likert (1932), la opción de respuesta de cinco puntos se ha popularizado y, junto con de siete opciones ha sido, por mucho, las más empleadas (Cummins, 1997). De hecho, el uso de la opción de respuesta de cinco puntos se ha convertido en una tradición, pero se requieren mayores argumentos para su uso (Bisquerra y Pérez-Escoda, 2015).

Con el propósito de verificar cuál opción de respuesta es mejor, han surgido estudios que han tratado de realizar diversas comparaciones. Aunque se ha abordado esta problemática, todavía no existe un consenso sobre cuál opción de respuesta es mejor (Lewis y Erdinç, 2017). Por un lado, ciertos autores no encontraron ventajas de alguna escala con algún número de respuestas en específico (Simms et al., 2019). Por ejemplo, Lewis y Erdinç (2017), luego de comparar las propiedades psicométricas (*i.e.*, confiabilidad y validez) de la escala Likert de 7 y 11 puntos, no encontraron diferencias significativas entre estas, concluyendo que dichos aspectos no parecen del todo relevantes. Mientras que Bisquerra y Pérez-Escoda (2015) recomiendan el uso de la escala de 11 puntos.



Otros estudios proponen que la opción de respuesta más recomendable es la de siete puntos (Finstad, 2010; Taherdoost, 2019), aunque esta valoración depende del objetivo de la investigación y la forma de administrar el cuestionario. Finalmente, están quienes mencionan que tanto la escala Likert de cinco y siete puntos no son adecuadas (Dolnicar et al., 2011), donde se pone mayor énfasis en la opción de respuesta de siete puntos, la cual consideran que puede presentar problemas de sesgo en las respuestas y, además, ser menos estable en comparación con una opción de respuesta dicotómica.

Al revisar los anteriores antecedentes de investigación, surgen preguntas como ¿Qué criterios de comparación han sido empleados en dichos estudios? Con el fin de verificar cuál opción de respuesta Likert es mejor, algunos estudios mencionan que tanto la validez como la confiabilidad son propiedades psicométricas muy importantes de una escala Likert (Cummins y Gullone, 2000). Otros estudios han evaluado la validez de las escalas (Chang, 1994; Hancock y Klockars, 1991; Preston y Colman, 2000). Bajo este contexto, la validez de un instrumento de medición puede ser concebida como el “el grado en que la evidencia empírica y los fundamentos teóricos respaldan la adecuación y pertinencia de las conclusiones extraídas de alguna forma de medida” (American Psychological Association [APA], 2015, p. 1129).

Considerando los objetivos de una investigación, existen diferentes tipos de validez (Aiken, 2003), debido a que su estudio es un proceso multifacético (Adcock y Collier, 2001). Primero, se encuentra la *validez de contenido* que mide qué tan bien los ítems representan un constructo (o al menos sus aspectos principales) de una variable latente (Hair et al., 2022), el cual, si corresponde a un instrumento en otro idioma, requiere de un proceso de traducción-retraducción (Lazić y Purić, 2022). Este tipo cuenta con fuentes de evidencia de validez. De acuerdo con APA (2015), dentro de las fuentes de evidencia de validez se destacan aquellas basadas en el contenido del test, en el proceso de respuesta, la estructura interna, su



relación con otras variables, y de las consecuencias de la interpretación de los resultados empleados. En este sentido, *la validez de constructo* es considerada una fuente de evidencia de validez que refleja el grado en que se mide una variable o factor (APA, 2010), la cual unifica la parte teórica con la evidencia empírica (Strauss y Smith, 2009). Además, junto a esta taxonomía, se encuentran la *validez convergente* (Sapsford, 2006), que corresponde al grado en que se asocian las medidas de un concepto (Hair et al., 2014), y la *validez discriminante* que, de acuerdo con Clark y Watson (2019), implica examinar cómo una medida se relaciona con indicadores de otros constructos, es decir, se buscan correlaciones de heterorrasgos.

Es importante señalar que el análisis de la validez ha generado la necesidad de reflexionar en mayor profundidad sobre el rol que juega la medición dentro de las ciencias sociales (Borsboom et al., 2004; Norman, 2016); partiendo de que la validez de un instrumento va más allá de la obtención de un coeficiente estadístico (Cronbach y Meehl, 1955); esta debe centrarse en la posibilidad de replicar un instrumento de medición en diferentes contextos (LeBel et al., 2017).

De esta forma, es importante considerar que otra de las ventajas de la escala Likert es la posibilidad de realizar estudios comparativos entre grupos (Hartley, 2014), en especial, entre diferentes culturas y países (Ariely y Davidov, 2012). Esto, luego de revisar que muchos instrumentos han sido estandarizados en países de habla inglesa, requiriendo un análisis especial para poder ser empleados en otros idiomas, lo cual no es del todo sencillo, pues puede presentar problemas en cuanto a la equivalencia cultural y/o semántica debido a la interpretación variada (Escobar-Pérez y Cuervo-Martínez, 2008). Es así que el análisis de invarianza, que ha sido empleada para medir la equivalencia —comparación de las respuestas en un conjunto de ítems entre grupos (Hair et al., 2022)— en estudios comparativos entre países (Ariely y Davidov, 2012), podría ser una alternativa metodológica para verificar qué escala de medición podría ser la más adecuada.



La invarianza hace alusión a que una medida sea semejante bajo diferentes condiciones de observación (Horn y Mcardle, 1992). De acuerdo con Hair et al. (2022), la invarianza involucra tres etapas: (1) *invarianza configural* (*i.e.*, igual parametrización y forma de estimación); (2) invarianza de composición (*i.e.*, ponderaciones iguales de los indicadores) e (3) igualdad de los valores medios compuestos y las varianzas. La invarianza de medida es entendida como la similitud o equivalencia en los resultados de un instrumento entre grupos (Byrne, 2008), lo cual juega un rol trascendental para validar a un instrumento de medición.

Es importante resaltar que, según Berrios-Riquelme et al. (2021), para evaluar la invarianza de un instrumento de medición se debe considerar lo siguiente: (a) *la invarianza configural*, donde se comprueba el ajuste del modelo sin ningún tipo de restricciones entre las muestras; (b) *la invarianza métrica*, es decir, si el ajuste continúa siendo adecuado aun cuando se encuentre una equivalencia de los modelos entre los grupos y, además, las cargas factoriales sean similares; (c) *la invariancia escalar*, se evalúa conservando las condiciones establecidas en los pasos anteriores —invarianza configural y métrica—, donde se restringen las intercepciones de los elementos con el fin de que los dos grupos sean equivalentes; una vez comprobándose esta invarianza, es posible completar el análisis a través de (d) *la invarianza estricta residual*, que establece una condición nueva que consiste en restringir las varianzas y covarianzas de los errores.

A partir de lo anterior, es posible observar que el estudio de la validez y la invarianza pueden servir como elementos de comparación para verificar qué escala es mejor, cinco o siete puntos. No obstante, la selección de una variable para realizar dicha contrastación requiere tres aspectos: (a) un cuestionario que haya sido validado en diferentes grupos (culturas o países); (b) que haya sido administrado en ambas escalas



(i.e., 5 y 7 puntos) y, a la vez, (c) que existan antecedentes de estudios sobre la invarianza de este en alguna de las escalas mencionadas.

Un instrumento que responde a dichas cualidades puede ser el cuestionario *Satisfaction with Life Scale* (SWLS), propuesto por Diener et al. (1985), el cual se encuentra integrado por cinco ítems que es respondido con siete opciones de respuesta tipo Likert. Es importante señalar que su evidencia empírica ha aumentado a nivel mundial (Pavot y Diener, 2008), lo cual ha reflejado resultados satisfactorios en cuanto a la validez de constructo y convergente (Espejo et al., 2022; Martell Muñoz et al., 2018; Pavot y Diener, 1993); y que ha sido medida mayormente a través de escalas Likert de siete y cinco puntos y, con menor frecuencia, de seis (Emerson et al., 2017).

Teóricamente, la satisfacción con la vida es concebida como la forma en que una persona evalúa de manera completa su vida de una forma positiva (Veenhoven, 2015). Asimismo, “es la medida en que una persona encuentra que la vida es rica, significativa, plena o de alta calidad” (APA, 2010, p. 456). Para Pavot y Diener (2009, p. 102), es “un juicio cognitivo consciente de la propia vida en el que los criterios de juicio dependen de la persona.” Su conceptualización puede variar entre las culturas (Oishi, 2006) y, además, existen críticas acerca de la validez de los instrumentos que miden a esta variable. Por ejemplo, Veenhoven (2015) considera que propiamente dichos instrumentos no la miden; esto puede deberse a lo siguiente: (a) la dificultad de replicar el instrumento en diferentes contextos; (b) la complejidad de controlar la parte cultural dentro de los estudios comparativos entre países (Vittersø et al., 2005); (c) la existencia de diferencias sobre cómo las personas conciben a la satisfacción con la vida (Tucker et al., 2006). Por lo tanto, hacer estudios comparativos entre grupos (p.ej. países, géneros) puede ayudar a probar equivalencias psicométricas entre



los ítems de una escala, lo cual es posible mediante la evaluación de la invarianza de medida (Hittner et al., 2018).

En la teoría de la medición es indispensable que exista un significado compartido entre los grupos donde la conceptualización de la variable que se desea medir sea la misma (Vittersø et al., 2005), ya que la escala puede no ser interpretada de la misma manera entre los grupos (Tucker et al., 2006). En esencia, al existir inconsistencias en cuanto a la invarianza de medida entre grupos, también habrá diferencias en las respuestas (Hittner et al., 2018).

Por otro lado, si bien es cierto que Diener et al. (1985) construyeron dicho instrumento mediante el empleo de escala Likert-7, existen otros estudios que han empleado otras escalas: seis puntos (Schnettler et al., 2017) y cinco puntos (Dimitrova y Domínguez, 2015; Padrós et al., 2015); evidencia empírica que reportó diferentes análisis de validez e invarianza entre grupos. Precisamente, bajo este último aspecto, se resalta que al ser la satisfacción con la vida una variable que puede cambiar según la parte cultural (Oishi, 2006), han surgido diferentes investigaciones que han realizado estudios de invarianza entre países (Jang et al., 2017; Jovanović et al., 2022).

Existen estudios que han investigado la validez e invarianza de la SWLS en diferentes países. Por ejemplo, Jang et al. (2017) evaluaron la validez de constructo y la invarianza configural a través de tres métodos diferentes en 15 idiomas y 26 países, entre ellos, países de habla hispana (Chile, Perú, Argentina, Bolivia, España) los autores encontraron invarianza configural y métrica, pero no comprobaron la invarianza escalar. Entre países de habla hispana existe evidencia empírica sobre validez de constructo del SWLS (Casas et al., 2012; Martell Muñoz et al., 2018; Mikulic et al., 2019; Padrós et al., 2015; Park et al., 2009) e



invarianza (Berrios-Riquelme *et al.*, 2021; Esnaola *et al.*, 2017; Espejo *et al.*, 2022; Jovanović *et al.*, 2022; Schnettler *et al.*, 2021).

En este mismo sentido, Jovanović *et al.* (2022), con una muestra más representativa en 19 idiomas y 24 países -de los cuales solo dos eran de habla hispana (Argentina y España)-, se midió la invarianza configural, métrica y escalar a través del Análisis Factorial Confirmatorio Multigrupo (AFC-Multigrupo) comparando a los países y continentes; donde los autores únicamente encontraron evidencia de invarianza configural entre países y continentes; los hallazgos encontrados no fueron concluyentes para las pruebas de invarianza métrica y escalar. Por otro lado, Park *et al.* (2009) aplicaron el instrumento de SWLS en 27 países; pero no analizaron la validez ni la invarianza del instrumento.

Como es posible observar, se ha buscado medir la invarianza a través de diversos países. La mayoría de sus muestras son de dos (Esnaola *et al.*, 2017; Sancho *et al.*, 2019; Schnettler *et al.*, 2017) o tres países (Berrios-Riquelme *et al.*, 2021; Dimitrova y Domínguez Espinosa, 2015; Whisman y Judd, 2016), pero son pocos los que estudian cuatro países (Casas *et al.*, 2012) o más (Jovanović y Brdar, 2018 [ver Tabla 1]). Es importante señalar que, según Emerson *et al.* (2017), la mayoría de los estudios que evalúan la invarianza de medida hacen uso del AFC-Multigrupo a través de la invarianza configural, métrica, escalar y estricta.



Tabla 1*Estudios empíricos sobre SWLS a nivel internacional y Latinoamérica*

Autor(es)	País(es) <i>n</i> = muestra	Muestra	Tipo de escala	Validez	Prueba estadística
Jovanović et al. (2022)	24 países Argentina España (<i>n</i> = 22 710)	Adolescentes	7	Invarianza (configural, métrica y escalar)	AFC-Multigrupo (comparación entre países y continentes)
Espejo et al. (2022)	Colombia (<i>n</i> = 1255)	Adultos	5	De constructo Invarianza (configural, métrica y escalar) Convergente Divergente Concurrente	AFC-Multigrupo (comparación entre género y edad) Correlación
Berrios- Riquelme et al. (2021)	Chile (<i>n</i> = 237) España (<i>n</i> = 242) EUA (<i>n</i> = 183)	Inmigrantes latinos	7	De constructo Invarianza (configural, métrica, escalar y residual)	AFC-Multigrupo (comparación entre países)
Schnettler et al. (2021)	Chile (<i>n</i> = 165) España (<i>n</i> = 109)	Estudiantes universidad	6	Invarianza (configural, métrica y escalar)	AFC-Multigrupo (comparación entre países y género)
Checa et al. (2019)	España (<i>n</i> = 726)	Adultos	7	Invarianza (configural, métrica y escalar)	AFC Multidimensional (comparación entre género, edad, estado civil y nivel educativo)
Sancho et al. (2019)	España (<i>n</i> = 857) Perú (<i>n</i> = 336)	Adultos mayores	5	De constructo Invarianza (configural, métrica, escalar y residual)	AFC-Multigrupo (comparación entre países, sexo y estado civil)
Esnaola et al. (2017)	España (<i>n</i> = 520) México (<i>n</i> = 181)	Adolescentes	7	De constructo Invarianza (configural, métrica, escalar y estricta)	AFC-Multigrupo (comparación entre países, género y edad)
Jang et al. (2017)	26 países Chile, Perú, Argentina, Bolivia, España (<i>n</i> = 7004)	Gerentes de empresas	7	De constructo Invarianza (configural, métrica y escalar)	AFC-Multigrupo (comparación entre países, género y edad) AFC-Multinivel Métodos de optimización de alineación
Schnettler et al. (2017)	Chile (<i>n</i> = 756) Ecuador (<i>n</i> = 817)	Adultos mayores	6	De constructo Convergente Invarianza (configural, métrica, escalar y estricta)	AFC AFC (comparación entre países) Correlaciones



Autor(es)	País(es) <i>n</i> = muestra	Muestra	Tipo de escala	Validez	Prueba estadística
Dimitrova y Dominguez (2015)	Argentina (<i>n</i> = 192) México (<i>n</i> = 421) Nicaragua (<i>n</i> = 302)	Adultos	5	De constructo Invarianza (configural, métrica y parcial)	AFC-Multigrupo (comparación entre países)
Casas et al. (2012)	Argentina (<i>n</i> = 342) Brasil (<i>n</i> = 1,588) Chile (<i>n</i> = 843) España (<i>n</i> = 2,900)	Adolescentes	100	De constructo Invarianza (métrica)	AFC-Multigrupo (comparación entre escalas y países)

Nota. EUA. = Estados Unidos de América; AFE = análisis factorial exploratorio; AFC = análisis factorial confirmatorio. Se mencionan solo los países iberoamericanos que formaron parte de las muestras estudiadas. Dichos estudios no se limitaron a los países de habla hispana.

En este contexto, surge la siguiente pregunta de investigación: ¿Cuál opción de respuesta Likert — cinco o siete puntos— es más apropiada para administrar el SWLS en países iberoamericanos mediante el análisis de invarianza? Como se puede observar, la evidencia empírica reportada no es del todo concluyente porque, si bien se ha reportado algún tipo de invarianza (configural, métrica, escalar o estricta), no se ha encontrado evidencia sobre la invarianza de medida de dicho instrumento, la cual requiere la comprobación de las cuatro pruebas antes mencionadas, y no solo de alguna de ellas. Por lo tanto, este estudio podría aportar lo siguiente: (a) unirse al debate psicométrico acerca de cuál opción de respuesta Likert puede ser mejor y (b) aportar evidencia empírica sobre el estudio de la invarianza de medida a través de la comparación entre dos o más grupos diferentes (*i.e.*, país y género).

Con base en los argumentos antes mencionados, la presente investigación se propuso comparar las propiedades psicométricas de ambas opciones de respuesta Likert, mediante el caso de un instrumento de satisfacción con la vida. Con la finalidad de comprobar si el número de opciones afecta la calidad métrica de la escala, se realizaron análisis estadísticos que permitieron comprobar la confiabilidad, validez e invarianza de medida del instrumento.



Metodología

Se realizó una investigación de tipo cuantitativa con un corte transversal y diseño no experimental. Se aplicó el cuestionario SWLS (Diener et al., 1985) a dos muestras de estudiantes universitarios de diferentes países iberoamericanos.

Participantes

Con el fin de obtener el mayor número de participantes en el estudio se empleó una muestra no probabilística por conveniencia invitando a colaborar en la investigación a jóvenes que radicaran en países de habla hispana. La muestra final se conformó por 1617 jóvenes hispanohablantes. A la primera muestra se le administró el cuestionario en físico con siete opciones de respuesta, en Argentina, Bolivia y Ecuador ($n_1 = 706$); mientras que, en el caso de la segunda muestra, se aplicó el mismo cuestionario con una escala de respuesta de cinco puntos en España, Bolivia y México ($n_2 = 911$). En la Tabla 2, se muestra la caracterización de los grupos estudiados.

Tabla 2

Caracterización de la muestra

	Grupo 1 ($n = 706$) Escala Likert-7		Grupo 2 ($n = 911$) Escala Likert-5		
	n	%	n	%	
Género			Género		
Femenino	466	66	Femenino	563	61.8
Masculino	240	34	Masculino	348	38.2
País			País		
Ecuador	162	22.9	México	416	45.7
Argentina	302	42.8	Bolivia	314	34.5
Bolivia	242	34.3	España	181	19.9
Edad	$M = 22.10$	$DE = 4.10$	Edad	$M = 20.84$	$DE = 2.7$



Instrumento de medición

Como se mencionó con anterioridad, se utilizó la escala propuesta por Diener et al. (1985) integrada por cinco ítems, la cual fue traducida al español y retraducida al idioma original. Para el grupo 1 se empleó una escala Likert-7 con los siguientes valores: 1 (“Totalmente en desacuerdo”), 2 (“En desacuerdo”), 3 (“Ligeramente en desacuerdo”), 4 (“Ni de acuerdo ni en desacuerdo”), 5 (“Ligeramente de acuerdo”), 6 (“De acuerdo”) y 7 (“Totalmente de acuerdo”). En cuanto al grupo 2, este tuvo una escala Likert-5, cuyos valores de medición fueron: 1 (“Totalmente en desacuerdo”), 2 (“En desacuerdo”), 3 (“Ni de acuerdo ni en desacuerdo”), 4 (“De acuerdo”) y 5 (“Totalmente de acuerdo”).

Análisis de datos

Para lograr la consecución del propósito del estudio, se realizaron los siguientes análisis: (1) confiabilidad a través del estadístico de alfa de Cronbach, cuyos valores aceptables son iguales o mayores a 0.70 (Hair et al., 2014); (2) análisis factorial exploratorio (AFE), considerando los siguientes indicadores: verificación de que sea una variable unifactorial, el porcentaje de la varianza explicada, la prueba Kaiser-Meyer-Olkin (KMO), cuyos valores aceptables son cercanos a 1.0 (Field, 2018); y prueba de esfericidad de Bartlett, donde se esperan valores p inferiores a 0.001 (Valdés et al., 2019). También se realizó un (3) análisis factorial confirmatorio (AFC), donde se evaluaron indicadores de bondad de ajuste estadístico (p -valor de X^2 , > 0.05), poblacional (RMSEA, < 0.08) y práctico (TLI, CFI, NFI, > 0.90 ; Hair et al., 2014, Valdés et al., 2019); (4) validez convergente a través del AVE, cuyos valores favorables son iguales o mayores a 0.50 (Chen et al., 2019) y (5) análisis de invarianza (configural, métrica, escalar y estricta), donde los indicadores de evaluación son los siguientes ΔCFI y $\Delta TLI \leq 0.010$, y $\Delta RMSEA < 0.015$ (Cheung y Rensvold, 2002; Jovanović et al., 2022). Cabe señalar que los dos primeros análisis fueron llevados a cabo en SPSS v. 23; mientras que el resto se llevaron a cabo en AMOS v. 27.



Procedimiento

Se solicitó el permiso de las autoridades de las universidades participantes para realizar la investigación, se obtuvo aprobación del Comité de Investigación Institucional y, además, los estudiantes participaron de manera libre y voluntaria dentro del presente estudio. Respondieron de forma física y digital cada una de las preguntas, teniendo la libertad de dejar de responder el cuestionario, se les aclaró que sus respuestas serían confidenciales y anónimas. Posterior a la recolección de información, se realizó la limpieza de las bases de datos y los análisis estadísticos antes mencionados.

Resultados

De manera preliminar, para explorar la validez de constructo, se realizó AFE. Se decidió utilizar un análisis de componentes principales con rotación varimax, debido a que rotar los factores permite mejorar las conjeturas de los análisis y maximizar la pureza de la saturación de las variables (Zeller, 2005). En la Tabla 3 se observan los resultados de tales análisis, los cuales denotan valores favorables.

Tabla 3

Análisis factorial exploratorio

Modelo	% de varianza explicada	KMO	Prueba de esfericidad de Bartlett
Modelo 1 (Likert-7)	58.072	0.836	$X^2 = 1134.01; p \leq 0.00$
Modelo 2 (Likert-5)	54.503	0.825	$X^2 = 1188.7; p \leq 0.00$

Para comprobar la validez de constructo, se llevó a cabo un AFC por medio de modelación con ecuaciones estructurales. Tal como se muestra en la Tabla 4, los índices de ajuste fueron aceptables para ambas muestras (TLI, CFI, NFI, RMSEA), excepto en el caso de X^2/gl , cuyo resultado fue significativo en ambas muestras. Cabe señalar que, según Ruiz et al. (2010), el valor aceptable es menor a 3.0, donde el estadístico de chi-cuadrado (X^2) debe ser no significativo; no obstante, este estadístico no es utilizado para



medir los índices de ajuste de los modelos debido a que es afectado por el tamaño de la muestra (Cheung y Rensvold, 2002; Putnick y Bornstein, 2016; Valdés et al., 2019).

Tabla 4

Análisis factorial confirmatorio

Modelos	X^2	gl	X^2/gl	p	IFI	TLI	CFI	NFI	RMSEA
Modelo 1 (Likert-7)	14.98	5	2.98	0.010	0.991	0.973	0.991	0.987	0.053
Modelo 2 (Likert-5)	16.15	15	3.23	0.006	0.991	0.972	0.991	0.986	0.050

Si bien los análisis anteriores (*i.e.*, AFE y AFC) son empleados para medir la validez de constructo, es importante resaltar que esta puede ser medida a través de la validez convergente, que se mide examinando las asociaciones entre ítems del mismo constructo, donde, a mayor nivel de convergencia, se podrá respaldar la validez del constructo (Clark y Watson, 2019). Esta se calcula por medio de la varianza extraída promedio (AVE). Además, se evaluó la consistencia interna (confiabilidad) de la escala, donde ambas muestras obtuvieron valores superiores a 0.70 en la prueba de alfa de Cronbach (ver Tabla 5).

Tabla 5

Validez convergente

Muestras	Escala	AVE	Confiabilidad (alfa de Cronbach)
Grupo 1	Likert 7	0.58	0.711
Grupo 2	Likert 5	0.54	0.731

Los hallazgos anteriores muestran la pertinencia para realizar un análisis de invarianza, en sus diferentes etapas: configural, métrica, escalar y estricta (residual). En primer lugar, en la invarianza configural se cumplió en ambas muestras al comparar el género, pero esto no sucedió en cuanto a los países (véase Tablas 6 y 7). En referencia al grupo 1, aunque obtuvo un resultado significativo, el indicador χ^2/gl fue superior a 3.0. Es importante poner atención en los valores de χ^2/gl , pues como señala Kline (2016), aunque los indicadores máximos de este parámetro no son claros, según Ruiz et al. (2010), este debería ser



inferior a 3.0, lo cual no sucedió en el caso de la invarianza estricta en el modelo de países y en el modelo de género en la invarianza configural, respectivamente.

Debido a este y otros aspectos, se recomiendan, como alternativas al χ^2/gl , índices como el GFI, TLI y RMSEA, donde los cambios (Δ) para los dos primeros indicadores deben ser menores a 0.010; mientras que, para el último, debe ser inferior a 0.015 (Chen, 2007; Cheung y Rensvold, 2002; Jovanović et al., 2022). Así, se identificó la invarianza métrica en el modelo de género del grupo 1.

Tabla 6

Invarianza de medición entre países y género del grupo 1 (escala Likert-7 puntos)

Modelo de países	X^2	gl	ΔX^2	p	X^2/gl	RMSEA	CFI	TLI	$\Delta RMSEA$	ΔCFI	ΔTLI
Configural	19.55	15		0.189	1.304	0.021	0.996	0.987			
Métrica	23.785	23	4.235	0.836	1.034	0.007	0.999	0.999	0.014	- 0.003	- 0.012
Escalar	98.941	33	75.156	0	2.998	0.053	0.940	0.918	- 0.046	0.059	0.081
Estricta	154.504	45	55.563	0	3.433	0.059	0.900	0.900	- 0.006	0.040	0.018
Modelo de género	X^2	gl	ΔX^2	p	X^2/gl	RMSEA	CFI	TLI	$\Delta RMSEA$	ΔCFI	ΔTLI
Configural	35.744	10		0	3.547	0.060	0.977	0.931			
Métrica	39.511	14	3.767	0.438	2.822	0.051	0.977	0.951	0.009	0	- 0.020
Escalar	49.019	19	9.508	0.151	2.48	0.047	0.973	0.957	0.004	0.004	- 0.006
Estricta	56.115	25	7.096	0.152	2.25	0.042	0.972	0.966	0.005	0.001	- 0.009

Nota. X^2 = chi-cuadrado; gl = grados de libertad; RMSEA = raíz del error medio cuadrático ajustado; CFI = índice comparativo de ajuste; TLI = índice Tucker-Lewis.

En cuanto al grupo 2, obtuvo resultados favorables en la invarianza configural tanto para los países como el género. Posteriormente, al evaluar la invarianza métrica, solo se identificaron indicadores adecuados en género (ΔCFI y $\Delta TLI \leq 0.010$, y $\Delta RMSEA < 0.015$). Finalmente, fue posible comprobar la invarianza escalar y estricta en el modelo de género (ver Tabla 7).



Tabla 7*Invarianza de medición entre países y género del grupo 2 (escala Likert-5 puntos)*

Modelo de países	X^2	gl	ΔX^2	p	X^2/gl	RMSEA	CFI	TLI	$\Delta RMSEA$	ΔCFI	ΔTLI
Configural	33.895	15		0.004	2.26	0.037	0.983	0.966			
Métrica	59.928	23	26.033	0.001	2.606	0.042	0.967	0.957	- 0.005	0.016	0.009
Escalar	177.025	33	117.097	0	5.364	0.069	0.872	0.883	- 0.027	0.095	0.074
Estricta	225.664	45	48.639	0	5.01	0.066	0.839	0.893	0.003	0.033	-0.010
Modelo de género	X^2	gl	ΔX^2	p	X^2/gl	RMSEA	CFI	TLI	$\Delta RMSEA$	ΔCFI	ΔTLI
Configural	20.09	10		0	2.009	0.033	0.991	0.983			
Métrica	22.732	14	3.767	0.619	1.624	0.026	0.993	0.989	0.007	- 0.002	-0.006
Escalar	36.9640	19	14.232	0.056	1.928	0.032	0.985	0.984	- 0.006	0.008	0.005
Estricta	41.674	25	4.898	0.078	1.674	0.042	0.986	0.989	- 0.010	- 0.001	-0.005

Nota. X^2 = chi-cuadrado; gl = grados de libertad; RMSEA = raíz del error medio cuadrático ajustado; CFI = índice comparativo de ajuste; TLI = índice Tucker-Lewis.

Discusión

Los hallazgos aportan evidencia empírica a estudios previos de invarianza de medida del SWLS, como los reportados por Emerson *et al.* (2017), quienes, luego de analizar 27 artículos, encontraron que la invarianza configural y métrica fueron las más frecuentes, siendo menos recurrentes la escalar y la estricta, en especial al usar al género como modelo de comparación. En cambio, en el caso de las culturas (países), es menos usual encontrar algún tipo de invarianza. Cabe señalar que los hallazgos encontrados en el presente estudio coinciden con lo reportado por los autores mencionados.

Entonces, ¿cuál opción de respuesta Likert es mejor? Si bien es cierto que los estudios previos no se han orientado en abordar esta cuestión del todo, el SWLS ha sido medido a través de diferentes escalas: siete (Arrindell *et al.*, 2022; Jovanović *et al.*, 2022), cinco (Dimitrova y Domínguez, 2015; Sancho *et al.*, 2019) y seis puntos (Schnettler *et al.*, 2017; Schnettler *et al.* 2021), e incluso, de 100 (Casas *et al.*, 2012); donde, también, se han realizado análisis de invarianza (ver Tabla 1), cuyos hallazgos han sido diversos. En el caso de aquellas investigaciones que usaron la escala Likert de siete opciones de respuesta, los resultados, aunque



no fueron concluyentes, sí reflejaron algún tipo de invarianza (Jovanović et al., 2022, respecto de invarianza configural en países y continentes; Jang et al., 2017, en invarianza configural y métrica). Por otro lado, respecto a la escala Likert de cinco opciones de respuesta, al igual que en la escala anterior, se identificó algún tipo de invarianza (Dimitrova y Domínguez, 2015, en invarianza configural, métrica y parcial al comparar muestras de Perú y España; Hultell y Gustavsson, 2008, en invarianza métrica entre sexo y edad, así como invarianza escalar entre sexos). Dichos resultados pueden no ser concluyentes, debido a que, de acuerdo con Hair et al. (2022), para evaluar la invarianza se requieren seguir tres pasos jerárquicamente interrelacionados; es decir, primero se debe comprobar la existencia de invarianza configural, luego, si se cumple con los parámetros, será posible evaluar el siguiente tipo de invarianza (métrica, escalar y estricta).

Al respecto, Putnick y Bornstein (2016) señalan que hace falta un consenso para poder comprobar la invarianza, generando con ello que muchos estudios de invarianza de medición reporten invarianza parcial para uno o más de los pasos, lo cual sugiere que los investigadores están adoptando prácticas de flexibilización al liberar restricciones, a fin de comprobar la no invarianza entre los grupos; no obstante, se conoce poco de las implicaciones conceptuales y estadísticas de tomar la invarianza parcial como válida. En el caso del SWLS, y el uso de la invarianza parcial, Emerson et al. (2017) señalan que algunos autores cuestionan la utilidad que tiene reportar invarianza parcial, debido a que si se modifican los subconjuntos de una escala (eliminar ítems), se alterará notablemente el constructo que se pretende evaluar; por lo tanto, sugieren que, si se pretende reportar una invarianza parcial, esta deberá estar teóricamente sustentada. En el caso de la presente investigación, se decidió no reportar una invarianza parcial.

Tomando en cuenta los argumentos expuestos, ¿es invariante el SWLS en cinco países hispanohablantes? En primer lugar, los resultados del presente trabajo no son concluyentes, pues no todos



los análisis de invarianza de medida han podido ser comprobados, al igual que varios de los estudios previos analizados por Emerson *et al.* (2017) y otros en población hispanohablante (Espejo *et al.*, 2022; Schnettler *et al.*, 2017; Schnettler *et al.*, 2021). Sin embargo, la evidencia empírica encontrada en este estudio sugiere que ambas escalas poseen invarianza configural y métrica entre géneros, pero no ha sido posible comprobar la invarianza escalar y estricta, excepto en el modelo de género de la escala Likert de cinco opciones de respuesta.

En cuanto a los hallazgos, estos implican que, si bien el instrumento es ampliamente utilizado, se requiere seguir realizando pruebas de confiabilidad y validez, debido a que los resultados pueden cambiar dependiendo del contexto; además, la aplicación en varios países puede aportar a la discusión sobre este tema. Así también, debería considerarse adaptar la escala al contexto actual, ya que esta fue propuesta en 1985. Por último, el presente estudio permitió evaluar su implicación metodológica con relación a la verificación de qué opción de respuesta puede ser la más adecuada.

Conclusiones

La presente investigación abordó múltiples propósitos: (a) determinar cuál escala Likert (cinco o siete puntos) puede ser mejor para medir el SWLS y (b) determinar la no invarianza de las escalas a través de países y el género en Iberoamérica. En primer lugar, ¿Cuál escala Likert es mejor? Si bien los hallazgos de este trabajo no fueron concluyentes, es posible sugerir, con base en estudios previos, que la escala Likert-5 obtuvo resultados más favorables.

En cuanto a la invarianza de medida, este trabajo contribuye al debate existente sobre la no invarianza de la escala SWLS, específicamente en cinco países hispanohablantes. Por lo tanto, se resalta que el



instrumento sí es invariante al comparar países, pero respecto al género se obtuvieron algunos indicadores favorables para soportar la invarianza del instrumento.

Considerando las limitaciones de la presente investigación, para próximos estudios sería recomendable atender los siguientes líneas de acción: (a) utilizar un muestreo probabilístico; (b) incluir en el estudio datos de otros países hispanohablantes; (c) realizar un comparativo con escalas Likert no sólo de cinco y siete puntos, sino también de seis, escala que podría dar continuidad a la discusión —ya que ha sido en menor medida empleada por otros autores (Schnettler et al., 2017; Schnettler et al., 2021)—, pero que, debido a que no se buscaba saturar a los participantes ni modificar el número de muestra, no se consideró en la presente investigación; (d) analizar la invarianza mediante AFC multigrupo con base en otros grupos de comparación (rango de edad, ingresos, grupos culturales); (e) medir la validez discriminante y (f) realizar otros análisis de la invarianza (de factores).

Conflicto de intereses

Los autores declaran no tener conflictos de intereses.

Financiamiento

Se agradece al Programa para el Fortalecimiento y Apoyo a la Investigación (PROFAPI, 2022) del Instituto Tecnológico de Sonora por el apoyo financiero recibido para la realización de esta investigación.

Agradecimientos

Se agradece a los participantes que colaboraron en responder el cuestionario de este estudio.



Referencias

- Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review*, 95(3), 529-546. <https://doi.org/10.1017/S0003055401003100>
- Aiken, L. R. (2003). *Test psicológicos y evaluación*. Peason Prentice Hall.
- Anjaria, K. (2022). Knowledge Derivation from Likert Scale Using Z-numbers. *Information Sciences*, 590, 234-252. <https://doi.org/10.1016/j.ins.2022.01.024>
- Ariely, G., & Davidov, E. (2012). Assessment of Measurement Equivalence with Cross-National and Longitudinal Surveys in Political Science. *European Political Science*, 11(3), 363-377. <https://doi.org/10.1057/eps.2011.11>
- Arrindell, W. A., Checa, I., Espejo, B., Chen, I.-H., Carrozzino, D., Vu-Bich, P., Dambach, H., & Vagos, P. (2022). Measurement Invariance and Construct Validity of the Satisfaction With Life Scale (SWLS) in Community Volunteers in Vietnam. *International Journal of Environmental Research and Public Health*, 19(6), 3460. <https://doi.org/10.3390/ijerph19063460>
- Asociación Americana de Psicología. (APA, 2010). *Diccionario conciso de psicología*. Manual Moderno.
- American Psychological Association. (APA, 2015). *APA Dictionary of Psychology* (2nd ed.). American Psychological Association.
- Berrios-Riquelme, J., Pascual-Soler, M., Frias-Navarro, D., & Maluenda-Albornoz, J. (2021). Psychometric Properties and Factorial Invariance of the Satisfaction with Life Scale in Latino Immigrants in Chile, Spain, and United States. *Terapia Psicológica*, 39(2), 199-218. <https://doi.org/10.4067/s0718-48082021000200199>
- Bisquerra, R. y Pérez-Escoda, N. (2015). ¿Pueden las escalas Likert aumentar en sensibilidad? *REIRE: Revista d'innovació i Recerca En Educació*, 8(2), 129-147. <https://doi.org/10.1344/reire2015.8.2828>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061-1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Bryman, A. (2003). *Quantity and Quality in Social Research*. Routledge. <https://doi.org/10.4324/9780203410028>
- Byrne, B. (2008). Testing for Multigroup Equivalence of a Measuring Instrument: A Walk Through the Process. *Psicothema*, 20(4), 872-882. <https://www.psicothema.com/pdf/3569.pdf>



- Carifio, J., & Perla, R. (2008). Resolving the 50-Year Debate Around Using and Misusing Likert Scales. *Medical Education*, 42(12), 1150-1152. <https://doi.org/10.1111/j.1365-2923.2008.03172.x>
- Casas, F., Sarriera, J. C., Abs, D., Coenders, G., Alfaro, J., Saforcada, E., & Tonon, G. (2012). Subjective Indicators of Personal Well-Being among Adolescents. Performance and Results for Different Scales in Latin-Language Speaking Countries: A Contribution to the International Debate. *Child Indicators Research*, 5(1), 1-28. <https://doi.org/10.1007/s12187-011-9119-1>
- 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. <https://doi.org/10.1177/014662169401800302>
- Checa, I., Perales, J., & Espejo, B. (2019). Measurement Invariance of the Satisfaction with Life Scale by Gender, Age, Marital Status and Educational Level. *Quality of Life Research*, 28(4), 963-968. <https://doi.org/10.1007/s11136-018-2066-2>
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Chen, X., Yu, Q., Yu, F., Huang, Y., & Zhang, L. (2019). Psychometric Evaluation of the Chinese Version of the Snizek-Revised Hall's Professionalism Inventory Scale. *Journal of International Medical Research*, 47(3), 1154-1168. <https://doi.org/10.1177/0300060518817401>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Clark, L. A., & Watson, D. (2019). Constructing Validity: New Developments in Creating Objective Measuring Instruments. *Psychological Assessment*, 31(12), 1412-1427. <https://doi.org/10.1037/pas0000626>
- Colman, A. M., Norris, C. E., & Preston, C. C. (1997). Comparing Rating Scales of Different Lengths: Equivalence of Scores from 5-Point and 7-Point Scales. *Psychological Reports*, 80(2), 355-362. <https://doi.org/10.2466/pr0.1997.80.2.355>
- Cox III, E. P. (1980). The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research*, 17(4), 407. <https://doi.org/10.2307/3150495>
- Croasmun, J. T., & 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>



- Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Cummins, R. A. (1997). *The Directory of Instruments to Measure Quality of Life and Cognate Areas of Study* (4th ed.). Deakin University.
- Cummins, R. A., & Gullone, E. (2000). *Why We Should Not Use 5-Point Likert Scales: The Case for Subjective Quality of Life Measurement* [conference]. Proceedings, Second International Conference on Quality of Life in Cities. National University of Singapore.
- Diener, E., Emmons, R. A., Larsen R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13
- Dimitrova, R., & Domínguez Espinosa, A. del C. (2015). Measurement Invariance of the Satisfaction with Life Scale in Argentina, Mexico and Nicaragua. *Social Inquiry into Well-Being*, 1(1), 32. <https://ri.iberomx/handle/iberomx/1063>
- Dolnicar, S., Grün, B., Leisch, F., & Rossiter, J. R. (2011, feb. 8). *Three Good Reasons NOT to Use Five and Seven Point Likert Items* [conference]. 21st CAUTHE National Conference. Adelaide, Australia.
- Emerson, S. D., Guhn, M., & Gadermann, A. M. (2017). Measurement Invariance of the Satisfaction with Life Scale: Reviewing Three Decades of Research. *Quality of Life Research*, 26(9), 2251-2264. <https://doi.org/10.1007/s11136-017-1552-2>
- Escobar-Pérez, J. y Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27-36. <https://doi.org/10.1016/j.edumed.2018.08.008>
- Esnaola, I., Benito, M., Antonio-Agirre, I., Freeman, J., & Sarasa, M. (2017). Measurement Invariance of the Satisfaction With Life Scale (SWLS) by Country, Gender and Age. *Psicothema*, 29(4), 596-601. <https://doi.org/10.7334/psicothema2016.394>
- Espejo, B., Martín-Carbonell, M., Checa, I., Paternina, Y., Fernández-Daza, M., Higueta, J. D., Albarracín, A., & Cerquera, A. (2022). Psychometric Properties of the Diener Satisfaction With Life Scale With Five Response Options Applied to the Colombian Population. *Frontiers in Public Health*, 9, 767534. <https://doi.org/10.3389/fpubh.2021.767534>
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Ltd.
- Finstad, K. (2010). Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales. *Usability Metric for User Experience*, 5(3), 104-110.



- <http://uxpajournal.org/response-interpolation-and-scale-sensitivity-evidence-against-5-point-scales/>
- Garwood, J. (2006). Likert Scale. In V. Jupp, *The SAGE Dictionary of Social Research Methods* (p. 161). SAGE Publications.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7^a ed.). Pearson Prentice-Hall.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE.
- Hancock, G. R., & Klockars, A. J. (1991). The Effect of Scale Manipulations on Validity: Targeting Frequency Rating Scales for Anticipated Performance Levels. *Applied Ergonomics*, 22(3), 147-154. <https://www.sciencedirect.com/journal/applied-ergonomics/vol/22/issue/3>
- Hartley, J. (2014). Some Thoughts on Likert-Type Scales. *International Journal of Clinical and Health Psychology*, 14(1), 83-86. [https://doi.org/10.1016/S1697-2600\(14\)70040-7](https://doi.org/10.1016/S1697-2600(14)70040-7)
- Heo, C. Y., Kim, B., Park, K., & Back, R. M. (2022). A Comparison of Best-Worst Scaling and Likert Scale Methods on Peer-to-Peer Accommodation Attributes. *Journal of Business Research*, 148, 368-377. <https://doi.org/10.1016/j.jbusres.2022.04.064>
- Hernández, R., Fernández, C. y Baptista, P. (2010). *Metodología de la investigación*. McGraw-Hill Interamericana.
- Hittner, J. B., Swickert, R., Silver, C. N., Hevesi, K., & Kövi, Z. (2018). Examining the Cross-National Measurement Invariance of the Extended Satisfaction with Life Scale in the United States and Hungary. *Applied Research in Quality of Life*, 13(3), 527-544. <https://doi.org/10.1007/s11482-017-9542-7>
- Horn, J. L., & Mcardle, J. J. (1992). A Practical and Theoretical Guide to Measurement Invariance in Aging Research. *Experimental Aging Research*, 18(3), 117-144. <https://doi.org/10.1080/03610739208253916>
- Hultell, D., & Gustavsson, J. P. (2008). A Psychometric Evaluation of the Satisfaction with Life Scale in a Swedish Nationwide Sample of University Students. *Personality and Individual Differences*, 44(5), 1070-1079. <https://doi.org/10.1016/j.paid.2007.10.030>
- Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., O'Driscoll, M. P., Sanchez, J. I., Spector, P. E., Poelmans, S. A. Y., Abarca, N., Alexandrova, M., Antoniou, A.-S., Beham, B., Brough, P., Carikci, I., Ferreira, P., Fraile, G., Geurts, S., ... Woo, J.-M. (2017). Measurement



- Invariance of the Satisfaction With Life Scale Across 26 Countries. *Journal of Cross-Cultural Psychology*, 48(4), 560-576. <https://doi.org/10.1177/0022022117697844>
- Jovanović, V., & Brdar, I. (2018). The Cross-National Measurement Invariance of the Satisfaction with Life Scale in A Sample of Undergraduate Students. *Personality and Individual Differences*, 128, 7-9. <https://doi.org/10.1016/j.paid.2018.02.010>
- Jovanović, V., Rudnev, M., Arslan, G., Buzea, C., Dimitrova, R., Góngora, V., Guse, T., Ho, R. T. H., Iqbal, N., Jámboři, S., Jhang, F.-H., Kaniušonytė, G., Li, J., Lim, Y.-J., Lodi, E., Mannerström, R., Marcionetti, J., Neto, F., Osin, E., ... Žukauskienė, R. (2022). The Satisfaction with Life Scale in Adolescent Samples: Measurement Invariance across 24 Countries and Regions, Age, and Gender. *Applied Research in Quality of Life*, 17(4), 2139-2161. <https://doi.org/10.1007/s11482-021-10024-w>
- Kerlinger, F. N. y Lee, H. B. (2002). *Investigación del comportamiento*. McGraw-Hill Interamericana.
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). The Guilford Press.
- Lazić, A., & Purić, D. (22-24 de septiembre de 2022). *Measurement Challenges in Multi-Site Projects: Translation and Adaptation of Psychological Instruments* [videoconference]. ABRIR Workshop Series “Increased Representation: A Vision for Inclusive Big-team Science”. Universidad de Sonora. <https://doi.org/10.17605/OSF.IO/28HMP>
- LeBel, E. P., Berger, D., Campbell, L., & Loving, T. (2017). Falsifiability is not optional. *Journal of Personality and Social Psychology*, 113(2), 254-261 <https://doi.org/10.1037/pspi0000106>
- Lewis, J. R., & Erdiñç, O. (2017). User Experience Rating Scales with 7,11, or 101 Points: Does It Matter? *Journal of Usability Studies*, 12(2), 73-91. <http://uxpajournal.org/user-experience-rating-scales-points/>
- Li, Q. (2013). A Novel Likert Scale Based on Fuzzy Sets Theory. *Expert Systems with Applications*, 40(5), 1609-1618. <https://doi.org/10.1016/j.eswa.2012.09.015>
- Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 22(140), 55-55.
- Martell Muñoz, J., García Cadena, C. H., Daniel-González, L., Sánchez Miranda, P. y Mendoza Martínez, A. (2018). Estructura factorial de la escala de satisfacción con la vida y validez convergente con la escala de calidad de vida en preparatorianos mexicanos. *Revista de Psicología y Ciencias del Comportamiento de la Unidad Académica de Ciencias Jurídicas y Sociales*, 9(2), 30-45. <https://doi.org/10.29365/rpcc.20181207-70>



- Mikulic, I. M., Crespi, M. y Caballero, R. Y. (2019). Escala de satisfacción con la vida (SWLS): estudio de las propiedades psicométricas en adultos de Buenos Aires. *Anuario de Investigaciones*, 51, 395-402. <https://www.redalyc.org/articulo.oa?id=369163433043>
- Norman, G. (2016). Is Psychometrics Science? *Advances in Health Sciences Education*, 21(4), 731-734. <https://doi.org/10.1007/s10459-016-9705-6>
- Oishi, S. (2006). The Concept of Life Satisfaction Across Cultures: An IRT Analysis. *Journal of Research in Personality*, 40(4), 411-423. <https://doi.org/10.1016/j.jrp.2005.02.002>
- Padrós, F., Gutiérrez, C. Y. y Medina, M. A. (2015). Propiedades Psicométricas de la escala de satisfacción con la vida (SWLS) de Diener en población de Michoacán (México). *Avances en Psicología Latinoamericana*, 33(2), 223-232. <https://doi.org/10.12804/apl33.02.2015.04>
- Park, N., Peterson, C., & Ruch, W. (2009). Orientations to Happiness and Life Satisfaction in Twenty-Seven Nations. *The Journal of Positive Psychology*, 4(4), 273-279. <https://doi.org/10.1080/17439760902933690>
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5(2), 164-172. <https://doi.org/10.1037/1040-3590.5.2.164>
- Pavot, W., & Diener, E. (2008). The Satisfaction With Life Scale and the Emerging Construct of Life Satisfaction. *The Journal of Positive Psychology*, 3(2), 137-152. <https://doi.org/10.1080/17439760701756946>
- Pavot, W., & Diener, E. (2009). Review of the Satisfaction With Life Scale. In E. Diener (Ed.), *Assessing Well-Being* (Vol. 39, pp. 101-117). Springer Netherlands. https://doi.org/10.1007/978-90-481-2354-4_5
- Preston, C. C. & Colman, A. M. (2000). Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences. *Acta Psychologica*, 104(1), 1-15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Ruiz, M. A., Pardo, A. y San Martín, R. (2010). Modelo de ecuaciones estructurales. *Papeles del Psicólogo*, 31(1), 34-45. <https://www.redalyc.org/articulo.oa?id=77812441004>



- Sancho, P., Caycho-Rodríguez, T., Ventura-León, J., Tomás, J. M., & Reyes-Bossio, M. (2019). Does the Spanish Version of the SWLS Measure the Same in Spain and Peru? *Revista Española de Geriátría y Gerontología*, 54(6), 329-333. <https://doi.org/10.1016/j.regg.2019.04.003>
- Sapsford, R. (2006). Validity of Measurement. In *The SAGE Dictionary of Social Research Methods* (pp. 314-315). SAGE Publications, Ltd. <https://doi.org/10.4135/9780857020116.n217>
- Schnettler, B., Miranda-Zapata, E., Lobos, G., del Carmen Lapo, M., Adasme-Berríos, C., & Hueche, C. (2017). Measurement Invariance in the Satisfaction with Life Scale in Chilean and Ecuadorian Older Adults. *Personality and Individual Differences*, 110, 96-101. <https://doi.org/10.1016/j.paid.2017.01.036>
- Schnettler, B., Miranda-Zapata, E., Sánchez, M., Orellana, L., Lobos, G., Adasme-Berríos, C., Sepúlveda, J., & Hueche, C. (2021). Cross-Cultural Measurement Invariance in the Satisfaction with Life Scale in Chilean and Spanish University Students. *Suma Psicológica*, 28(1). <https://doi.org/10.14349/sumapsi.2021.v28.n2.8>
- 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. <https://doi.org/10.1037/pas0000648>
- Singh, D. P., Ahmed, N., & Gupta, N. (2021). *Business Research Method and Project Work*. SBPD Publications.
- Strauss, M. E., & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology*, 5(1), 1-25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639>
- Taherdoost, H. (2019). What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale. *International Journal of Academic Research in Management*, 8(1), 1-10. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3588604
- Tucker, K. L., Ozer, D. J., Lyubomirsky, S., & Boehm, J. K. (2006). Testing for Measurement Invariance in the Satisfaction with Life Scale: A Comparison of Russians and North Americans. *Social Indicators Research*, 78(2), 341-360. <https://doi.org/10.1007/s11205-005-1037-5>
- Valdés, M. A., García, F. I., Torres, G. M., Urías, M., & Grijalva, C. S. (2019). *Medición en investigación educativa con apoyo del SPSS y el AMOS*. Clave Editorial.



- Veenhoven, R. (2015). The Overall Satisfaction with Life: Subjective Approaches (1). En W. Glatzer, L. Camfield, V. Møller y M. Rojas (Eds.), *Global Handbook of Quality of Life* (pp. 207-238). Springer Netherlands. https://doi.org/10.1007/978-94-017-9178-6_9
- Vittersø, J., Biswas-Diener, R., & Diener, E. (2005). The Divergent Meanings of Life Satisfaction: Item Response Modeling of the Satisfaction with Life Scale in Greenland and Norway. *Social Indicators Research*, 74(2), 327-348. <https://doi.org/10.1007/s11205-004-4644-7>
- Whisman, M. A., & Judd, C. M. (2016). A Cross-National Analysis of Measurement Invariance of the Satisfaction With Life Scale. *Psychological Assessment*, 28(2), 239-244. <https://doi.org/10.1037/pas0000181>
- Zeller, R. A. (2005). Measurement Error, Issues and Solutions. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 665-676). Elsevier. <https://doi.org/10.1016/B0-12-369398-5/00109-2>

Cómo citar este artículo: Núñez Ramírez, M. A., Garduño Realivazquez, K. A., & Esparza García, I. G. (2023). Invarianza en escala Likert de cinco y siete puntos del SWLS en 5 países iberoamericanos. *Psicumex*, 14(1), 1-29, e605. <https://doi.org/10.36793/psicumex.v14i1.605>

