



Nuñez-Ramírez, M. A.; Garduño-Realivazquez, K. A.; Velarde-Flores, C. L. (2023). Comparación de las escalas Likert de concordancia y frecuencia para medir la satisfacción con la vida en trabajadores universitarios. *Journal of Sport and Health Research.* 17(4):11-22. <https://doi.org/10.58727/jshr.102085>

Original

COMPARACIÓN DE LAS ESCALAS LIKERT DE CONCORDANCIA Y FRECUENCIA PARA MEDIR LA SATISFACCIÓN CON LA VIDA EN TRABAJADORES UNIVERSITARIOS

COMPARISON OF LIKERT SCALES OF AGREEMENT AND FREQUENCY TO MEASURE LIFE SATISFACTION IN UNIVERSITY WORKERS

Nuñez-Ramírez, M. A.¹; Garduño-Realivazquez, K. A.¹; Velarde-Flores, C. L.²

¹*Instituto Tecnológico de Sonora, Departamento de Ciencias Administrativas. Ciudad Obregón, Sonora (México)*

²*Universidad de Sonora, Departamento de Contabilidad. Hermosillo, Sonora (México).*

Correspondence to:
Karla Alejandra Garduño-Realivazquez
 Instituto Tecnológico de Sonora
 Departamento de Ciencias Administrativas
 5 de Febrero #818 sur
 Email: karla.realivazquez@gmail.com

Edited by: D.A.A. Scientific Section
 Martos (Spain)



Received: 01/10/2023
 Accepted: 11/10/2024



¿LA OPCIÓN DE RESPUESTA EN ESCALA LIKERT IMPORTA? COMPARACIÓN ENTRE LAS OPCIONES DE CONCORDANCIA Y FRECUENCIA AL MEDIR SATISFACCIÓN CON LA VIDA.

RESUMEN

La escala tipo Likert es una de las herramientas psicométricas más empleadas en la actualidad, sin embargo, pueden ser sensibles a los cambios de circunstancias. Por tal motivo, el objetivo fue comparar las opciones de respuesta tipo Likert de concordancia y de frecuencia de una escala que mide la satisfacción con la vida en trabajadores universitarios. Mediante una metodología cuantitativa de alcance descriptivo-comparativo, se recolectaron dos muestras de estudio en el estado de Sonora, México: una conformada por profesores ($n_1= 211$) y otra por administrativos ($n_2= 205$). Para ello, se realizó una comparación de medidas y análisis factorial confirmatorio, al primer grupo se le administró el instrumento con la escala original (de concordancia), mientras que, al segundo, una escala de frecuencia. Dentro de los resultados no se identificaron diferencias significativas entre las opciones de respuesta, siendo la escala de concordancia la que mostró mejores resultados. En conclusión, la versión de la escala Likert original del instrumento obtuvo mejores resultados en cuanto a los análisis de validez de constructo y de confiabilidad. Este tipo de análisis puede arrojar pautas para fortalecer la comprensión de la validez de constructo de un instrumento cuantitativo, en especial para medir la satisfacción con la vida.

Palabras clave: escala Likert; propiedades psicométricas; satisfacción con la vida; trabajadores universitarios.

DOES THE LIKERT SCALE RESPONSE OPTION MATTER? COMPARISON BETWEEN CONCORDANCE AND FREQUENCY OPTIONS WHEN MEASURING LIFE SATISFACTION.

ABSTRACT

The Likert scale is one of the most used psychometric tools nowadays, however, they can be sensitive to change of circumstances. For this reason, the objective was to compare the Likert-type response options of agreement and frequency of a scale that measures life satisfaction in university workers. Using quantitative methodology of a descriptive-comparative scope, two samples were collected in the state of Sonora, Mexico, including professors ($n_1= 211$) and administrators ($n_2= 205$). The first group responded to a survey with the original scale (agreement), while the second group was given a frequency scale. Within the results, no significant differences were identified between the response options, with the agreement scale showing the best results. In conclusion, the original Likert scale version of the instrument obtained better results in terms of construct validity and reliability analyses. This type of analysis provides guidelines to strengthen the understanding of the construct validity through a quantitative approach, especially to measure life satisfaction.

Keywords: *Likert scale; psychometric properties; Life satisfaction; university workers.*



INTRODUCCIÓN

Las escalas de medida suelen ser sensibles a los cambios de circunstancias de la vida de las personas (Diener et al., 2013). Por tal motivo, investigadores han buscado explorar las propiedades psicométricas de diversos constructos, principalmente a través de la validez de las mismas (Moreta-Herrera et al., 2023). Se entiende como validez al grado en que un constructo mide la realidad a través de manifestaciones de una variable latente (de Ayala, 2022). Para ello, se han empleado enfoques como la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI), con el objeto de evaluar la equivalencia de medida, la consistencia interna y la validez (Moreta-Herrera et al., 2023). Específicamente, la TRI es uno de los modelos probabilísticos más utilizados para medir las variables latentes (i.e., atributos que no son observados directamente); esta teoría es principalmente útil para examinar la dimensionalidad de los constructos con el objeto de determinar si la variable puede ser medida sin ambigüedades (Aryadoust & Raquel, 2019). La TRI representa un método estadístico que busca evaluar la probabilidad de que una persona logre obtener una puntuación específica cuando el contexto o la situación a la que se enfrenta cambia (Alimoradi et al., 2022). Cabe destacar que, uno de los tópicos que han llamado la atención en lo referente a la TRI, es el uso de las escalas Likert (Jebb et al., 2021).

Una de las principales herramientas psicométricas para realizar investigación cuantitativa, es la escala tipo Likert (Anjaria, 2022), la cual es empleada en diferentes campos de estudio como las ciencias sociales. Su origen se remonta a Freyd, en 1923; sin embargo, su popularidad creció con Rensis Likert, en 1932 (Bisquerra & Pérez-Escoda, 2015), a quien se le debe el nombre de esta escala. Ésta suele ser muy útil cuando se busca obtener una medición sobre un constructo, opinión o experiencia, así como para recopilar datos específicos sobre un tema (Roy, 2020). Además, con el paso del tiempo ha demostrado tener ciertas ventajas y desventajas (Hartley, 2014). En cuanto a lo primero, surge la pregunta, ¿por qué se ha hecho tan popular? Algunas de las ventajas al realizar investigación con esta escala se encuentran: (a) su uso es fácil de comprender (Anjaria, 2022); (b) sirve para medir datos de rasgos subjetivos (South et al., 2022); (c) permite la codificación de datos cuantitativos (Heo et al., 2022); (d) favorece la medición de variables no

observables (Jebb et al., 2021); y (e) puede ser empleada para comparar grupos (Hartley, 2014). Además, otra de las bondades de esta escala es que cuenta con una variedad de categorías para medir las variables, lo cual permite adecuar las opciones al objetivo del instrumento de medición. Éstas se catalogan en diferentes tipos: de concordancia (Totalmente de acuerdo a Totalmente en desacuerdo), frecuencia (Frecuentemente a Nunca), nivel de satisfacción (Muy satisfecho a Muy insatisfecho), importancia (Muy importante a No es importante), probabilidad (Extremadamente probable a No es probable), calidad (Muy buena a Muy mala), entre otras (Roy, 2020).

Sin embargo, las escalas Likert han mostrado ciertas deficiencias: (a) ambigüedad en el entendimiento de las palabras debido a la variedad de significados sobre un tópico, dado que algunas personas pueden tener problemas para diferenciar entre de acuerdo y totalmente de acuerdo o, en su caso, entre nunca y siempre (Mizumoto & Takeuchi, 2009); (b) los participantes pueden entender vagamente las diferentes categorías (e.g., Algunas veces o Frecuentemente; Brown, 2004); (c) puede existir ambigüedad en cuanto a la exactitud de los datos recopilados, la distancia entre las opciones del instrumento y los criterios para el uso de diferentes escalas (Anjaria, 2022); (d) sobresale un debate sobre cuál es el formato adecuado para medir las variables, ya que algunos investigadores confunden la diferencia de las escalas (South et al., 2022). La última desventaja, se debe a que algunas escalas pueden tener un adecuado desempeño bajo ciertas circunstancias, por ejemplo, la escala de frecuencia favorece la obtención de datos en rango normal y colectan información más precisa en cuanto a comportamientos, reacciones o situaciones; mientras que la escala de acuerdo es mejor para medir los valores extremos e identificar la percepción subjetiva sobre actitudes, comportamientos o sentimientos (Marfleo et al., 2014).

Han surgido estudios orientados hacia la comparación de las opciones de respuesta Likert con otras variables (e.g. evaluación funcional de discapacidad de trabajo [Marfleo et al., 2014], actitud con calificaciones de autoinforme [Brown, 2004]). Éstas, esencialmente, se pueden integrar en tres grupos. En primer lugar, están los que comparan las dos opciones de respuesta



mediante los resultados descriptivos; mientras que, en segundo lugar, se ubican aquellos estudios que ponen énfasis en realizar dicha comparación por medio de los resultados de la medición de validez de ambas escalas; y, finalmente, están los estudios que comparan las opciones de respuesta por su nivel predictivo.

En cuanto a la comparación de datos descriptivos, Brown (2004) contrastó las opciones de respuesta Likert de concordancia y de frecuencia de una escala orientada a medir la actitud con calificaciones de autoinforme agrupadas positivamente. Este autor reportó que la evidencia empírica no fue concluyente para afirmar que la opción de respuesta de concordancia es superior a la de frecuencia al medir esta variable. Mediante el uso de indicadores descriptivos (e.g., frecuencias, desviaciones, medidas, desviación estándar), se identificó que los resultados de ambas opciones de respuesta son similares; mientras que, al comparar los datos, se encontró que los resultados descriptivos no tuvieron grandes diferencias significativas, a la par que las correlaciones sí reflejaron diferencias, respectivamente. Además, Marfeo et al. (2014) trataron de investigar el óptimo funcionamiento de las opciones de respuesta de frecuencia y concordancia en un cuestionario desarrollado sobre evaluación funcional de discapacidad de trabajo, mediante escalas de estado de ánimo, emociones y control del comportamiento. Aquí se encontró que, en ambas opciones de respuesta, la confiabilidad fue adecuada; sin embargo, la de frecuencia mostró mejores resultados en la desviación estándar, mientras que la opción de respuesta de concordancia obtuvo mayores oscilaciones en las respuestas de los valores extremos. A partir de esto, los autores concluyeron que es necesario un formato que mezcle ambas opciones de respuesta para aprovechar las bondades de cada una.

En cuanto al segundo grupo, se destaca el estudio de Mizumoto y Takeuchi (2009), quienes compararon dos opciones de respuesta Likert de frecuencia y veracidad de un cuestionario sobre estrategias de aprendizaje de lenguaje mediante análisis factorial confirmatorio (AFC), encontrando que la opción de respuesta Likert de veracidad obtuvo resultados ligeramente mejores en cuanto a algunas medidas de bondad de ajuste del modelo (e.g., χ^2/gI , AGFI, CFI). Finalmente, Tong et al. (2020) compararon las opciones de respuesta de frecuencia y concordancia

para medir *burnout* y compromiso en dos muestras de trabajadores, cuyos principales hallazgos fueron que la escala de frecuencia presentó un mejor desempeño en ambas variables. Cabe señalar que dichos autores realizaron un AFC; sin embargo, no realizaron una comparación de los hallazgos.

Como se puede apreciar, a pesar de los avances en el estudio de las propiedades psicométricas de la escala Likert, se destacan las siguientes premisas: (a) todavía falta mayor evidencia empírica sobre las diferencias en el uso de las categorías al tratar de medir un constructo dentro de las ciencias sociales; (b) los estudios citados -los cuales han tratado este tópico- han hecho comparaciones entre las opciones de respuesta a través de las frecuencias o la validez de constructo; y, finalmente, (c) para poder contribuir a esta discusión -la cual, según Marfeo et al. (2014), es una temática poco estudiada con importantes implicaciones para la psicometría-, se pretende hacer un estudio que compare ambas opciones de respuesta (de frecuencia y concordancia), por medio de la estadística descriptiva, como lo reportan otros estudios (e.g., Brown, 2004; Marfeo et al., 2014), y, al mismo tiempo, a través del análisis de validez de constructo, como lo hizo Mizumoto y Takeuchi (2009).

Para tratar de integrarse a esta discusión se propone hacer una comparación de dos opciones de respuesta Likert con dos categorías diferentes: de frecuencia y concordancia. Con este propósito se pretende realizar dicho análisis en un instrumento ampliamente validado para medir la satisfacción con la vida denominado en inglés como *Satisfaction With Life Scale* (SWLS), el cual fue propuesto por Diener et al. (1985), quienes conciben a esta variable como un juicio cognitivo a través del cual una persona determina qué tan satisfecha con su vida está al comparar sus propias circunstancias con un modelo a seguir. Cabe señalar que, aunque han surgido diferentes escalas para medir esta variable, las cuales han mostrado niveles favorable de validez y confiabilidad (Diener et al., 2013), el SWLS es un instrumento que ha sido ampliamente replicado a lo largo del mundo, motivo por el cual la investigación en cuanto a su validez de constructo ha ido más allá y, en los últimos años, se ha puesto central énfasis en la medición de la invarianza de medida -en especial a través de AFC multigrupo-, calidad que refleja un nivel más exhaustivo de validez.



Por ejemplo, Moreta-Herrera et al. (2023) se propusieron verificar la equivalencia de medida del instrumento SWLS, a través de la TCT y TRI. Los autores, mediante los criterios de TCT, encontraron validez métrica y consistencia interna, en cambio con los criterios de la TRI identificaron que las respuestas de los ítems cambian dependiendo del nivel de habilidad de las personas; es importante mencionar que estos utilizaron una escala Likert de concordancia de 7 puntos. Por su parte, Jang et al. (2017) y Emerson et al. (2017) revisaron diferentes estudios que se han centrado en la invarianza de medida, encontrando que se ha avanzado de manera relevante en este tópico; no obstante, la evidencia empírica no es concluyente si el SWLS cumple con todos los criterios para considerarse como invariante. Esto da pauta a otro tipo de estudios que ahonden en la validez de dicho instrumento, considerando caminos alternos como la comparación de escalas, como puede ser la escala de frecuencia y la de concordancia. Cabe destacar que esta última es la escala propuesta por Diener et al. (1985), en la cual se encuentra en la mayoría de la evidencia empírica. Entonces, respecto al estudio de la validez del SWLS, ¿qué se podría aprender al comparar las escalas Likert?

Lo anterior se sustenta en las siguientes premisas: (a) la validez, como calidad psicométrica de un instrumento, es un concepto unitario -es decir, solo hay un tipo de validez-, que puede ser medida en diferentes procedimientos (Jebb et al., 2021); (b) las escalas orientadas a medir variables de salud, como puede ser la SWLS, pueden ser subjetivas (Marfeo et al., 2014), aspecto que requiere mayor investigación; y, finalmente (c) si bien es cierto que el SWLS ha sido aplicado dentro del contexto universitario, su uso ha estado sesgado hacia estudiantes (e.g., Jovanović & Brdar, 2018) y profesores (Emerson et al., 2017), mientras que la evidencia empírica es menor con trabajadores administrativos.

Al respecto, González-Rico et al. (2018) afirman que los trabajadores universitarios han sido marginados al explorar el bienestar de estos en su entorno laboral, así como en sus vidas personales fuera del lugar de trabajo. Por tal motivo, estos autores y González-Rico et al. (2022) analizaron factores que generan bienestar -como lo es satisfacción con la vida- en personal académico de investigación, y personal de servicio y administrativo; específicamente la SWLS reportó

adecuados indicadores de confiabilidad $\geq .88$. Cabe mencionar que dichos autores no reportaron indicadores de validez de la escala. Por lo tanto, dentro de la presente investigación, se pretende estudiar a trabajadores universitarios de México, incluyendo tanto académicos como administrativos.

Por otro lado, el realizar una comparación de dichas escalas, tanto a nivel de datos descriptivos como de validez, podría, al mismo tiempo, dar continuidad en el estudio sobre las propiedades psicométricas de la escala Likert, tal como lo han hecho con otras variables ya mencionadas, para poder aportar a la discusión sobre la validez del constructo del SWLS, donde se han dado pasos importantes, según las revisiones hechas por Jang et al. (2017) y Emerson et al. (2017). A partir de lo anterior, surge el siguiente objetivo de investigación: comparar las opciones de respuesta tipo Likert de concordancia y de frecuencia de una escala que mide la satisfacción con la vida en trabajadores universitarios.

MATERIAL Y MÉTODOS

Con el fin de cumplir el objetivo del estudio, que consiste en comparar dos categorías de la escala SWLS, se realizó un estudio comparativo descriptivo transversal de tipo cuantitativo.

Sujetos

A través de un muestreo no probabilístico intencional, se obtuvo una muestra total de 416 de trabajadores de universidades del noroeste de México, la cual se dividió en dos grupos: (1) profesores ($n_1 = 211$) y (2) administrativos ($n_2 = 205$). La caracterización de estos se muestra en la Tabla 1.

Instrumentos de medición

Se empleó una versión en español de la escala SWLS propuesta por Diener et al. (1985), compuesta por cinco ítems en escala Likert de concordancia de siete puntos. Estudios han reportado niveles adecuados de confiabilidad y validez de constructo en países de habla hispana (Espejo et al., 2022; Mikulic et al., 2019). Se utilizaron dos opciones de respuesta Likert para medir la satisfacción con la vida en trabajadores universitarios, donde dependiendo del grupo de estudio se les asignó las siguientes opciones de respuesta: (a) opción de respuesta de concordancia: 1(Totalmente en desacuerdo), 2(Muy en desacuerdo), 3(En desacuerdo), 4(Ni de acuerdo ni en desacuerdo), 5(De acuerdo), 6(Muy de acuerdo) y 7(Totalmente de acuerdo).



acuerdo); (b) opción de respuesta de frecuencia: 1(Nunca), 2(Muy rara vez), 3(En algunas ocasiones), 4(En bastantes ocasiones), 5(Con frecuencia), 6(Casi siempre) y 7(Siempre).

Procedimiento

Para realizar la investigación se obtuvo la autorización por parte del Comité de Ética del Instituto Tecnológico de Sonora para llevar a cabo el estudio. El instrumento fue administrado de forma presencial y en modalidad virtual (electrónica). Los participantes, luego de haber leído el consentimiento informado, decidieron participar libremente en la investigación. Cabe destacar que solo se tuvo un criterio de inclusión para formar parte del estudio: ser trabajadores universitarios en el estado de Sonora. Se realizaron dos análisis estadísticos: (a) análisis descriptivo-comparativo, mediante comparación de medias para datos no paramétricos (U de Mann-Whitney, el cual es

recomendado la comparación de medidas de muestras no paramétricas, así como cuando se busca evaluar ítems de manera independiente [Bertram, 2007]); (b) análisis factorial exploratorio (AFE) y AFC para ambas muestras, pruebas a las que ha respondido adecuadamente el SWLS en diferentes estudios (e.g., Jovanović et al., 2022); para ello, se emplearon los programas SPSS v. 27 y AMOS v. 29, respectivamente. En cuanto al primero, se consideraron los indicadores de Kaiser Meyer-Olkin (KMO), la Prueba de Esfericidad de Bartlett, la varianza total explicada y la agrupación de los ítems. Mientras que, en cuanto al segundo, se tomaron como referencia los índices de bondad de ajuste del modelo estructural, así como la AVE (varianza promedio extraída), esta última para medir la validez convergente. En ambas pruebas se trató de medir el nivel de validez de constructo del instrumento, así como la confiabilidad.

Tabla 1. Caracterización de la muestra

	Profesores (n ₁ = 211)		Administrativos (n ₂ = 205)	
	n	%	n	%
Sexo				
Femenino	120	56.9	151	73.7
Masculino	91	43.1	54	26.3
Edad	M = 43.15	DE= 10.5	M = 35.7	DE = 8.1
Antigüedad	M = 10.92	DE = 9.3	M = 9.5	DE = 7.6
Escolaridad				
Bachillerato	1	0.5	33	16.1
Especialidad	1	0.5	59	28.8
Licenciatura	39	18.5	103	50.2
Maestría	105	49.8	9	4.4
Doctorado	65	30.8	1	0.5

RESULTADOS

Los descriptivos de ambas opciones de respuesta muestran que: (a) los porcentajes se agruparon mayormente en los ítems más altos (5, 6 y 7); (b) respecto a la normalidad, considerando la asimetría y curtosis, se encontró que todos los ítems de las dos

opciones de respuesta tuvieron una distribución normal con excepción de los ítems 3 y 4 del grupo 1 con curtosis superior a 2. De acuerdo con Sürümü et al. (2023), los valores entre ± 1 son excelentes; sin embargo, estadísticos entre ± 2 son considerados aceptables (ver Tablas 2 y 3).

Tabla 2. Porcentajes de las Frecuencias Grupo 1 (n₁ = 211)



Ítems	$n_1 = 211$							Asimetría		Curtosis	
	1	2	3	4	5	6	7	Estadístico	S.E.	Estadístico	S.E.
Ítem 1	0.5	3.3	7.6	7.6	14.2	37	29.9	-1.111	0.167	0.543	0.333
Ítem 2	0.9	2.4	5.7	12.8	19.0	38.4	20.9	-0.975	0.167	0.675	0.333
Ítem 3	0.0	0.9	2.4	5.2	11.4	34.1	46.0	-1.485	0.167	2.235	0.333
Ítem 4	0.5	0.5	4.3	2.8	13.7	36.5	41.7	-1.590	0.167	2.966	0.333
Ítem 5	0.9	1.9	5.7	8.5	18.0	28.9	36.0	-1.131	0.167	0.883	0.333

Tabla 3. Porcentajes de las Frecuencias Grupo 2 ($n_2 = 205$)

Ítems	$n_2 = 205$							Asimetría		Curtosis	
	1	2	3	4	5	6	7	Estadístico	S.E.	Estadístico	S.E.
Ítem 1	11.2	13.2	21	13.7	15.6	16.1	9.3	0.032	0.170	-1.102	0.338
Ítem 2	1.0	3.9	11.2	17.1	16.6	30.7	19.5	-0.572	0.170	-0.520	0.338
Ítem 3	0.5	2	7.8	11.2	13.2	26.8	38.5	-0.966	0.170	0.050	0.338
Ítem 4	1.5	2.9	9.3	10.2	15.1	36.6	24.4	-0.973	0.170	0.267	0.338
Ítem 5	12.2	10.7	20	7.3	12.2.	20.5	17.1	-0.152	0.170	-1.329	0.338

Ahora bien, con el fin de cumplir con el objetivo de la investigación, en un primer momento, se realizó la comparación de medidas de ambas muestras. Para ello se utilizó la prueba de U de Mann-Whitney, la cual es útil para determinar qué grupo obtiene puntuaciones más altas, y que dichas diferencias sean significativas (Field, 2018). Los hallazgos de la prueba U de Mann-

Whitney indican que, aunque la muestra donde se aplicó la escala Likert de concordancia ($n_1 = 211$) obtuvieron medias (M) mayores, con desviaciones estándar (DE) más cercanas al valor 1.0 en todos los casos; no obstante, no se encontraron diferencias significativas (ver Tabla 4).

Tabla 4. Comparación de medias

Ítems	$n_1 = 211$		$n_2 = 205$		U de Mann-Whitney
	M	DE	M	DE	
Ítem 1	5.62	1.39	3.95	1.84	No significativo
Ítem 2	5.45	1.32	5.15	1.48	No significativo
Ítem 3	6.13	1.06	5.69	1.42	No significativo
Ítem 4	6.05	1.11	5.42	1.46	No significativo
Ítem 5	5.71	1.36	4.26	2.03	No significativo

Por otro lado, luego de haber realizado análisis de confiabilidad, se buscó comprobar la estructura factorial del modelo, para ello se realizó un AFE utilizando el análisis de componentes principales, con método de rotación Varimax con normalización Kaiser; en la Tabla 5 se presentan los resultados obtenidos en el primer análisis

factorial. Se consideraron los indicadores de KMO y alfa de Cronbach, cuyos valores mínimos aceptables son de $\geq .70$, además de la prueba de esfericidad de Bartlett, la cual debe ser significativa (Field, 2018; Hair et al., 2022). Los indicadores de confiabilidad y de análisis de componentes principales resultaron mejores para



la opción de respuesta de concordancia del SWLS, que es la escala original (e.g., Diener et al., 2013; 1985). Cabe destacar que la opción de respuesta de concordancia ha sido estudiada en

diversos contextos del mundo y tiende a reportar coeficientes de alfa $\geq .80$ (González-Rico et al., 2022; Moreta-Herrera et al., 2023).

Tabla 5. Análisis Factorial Exploratorio

Indicador	Concordancia ($n_1 = 211$)	Frecuencia ($n_2 = 205$)
Confiabilidad	$\alpha = .871$	$\alpha = .592$
Kaiser Meyer Olkin	KMO = .847	KMO = .720
Prueba de Esfericidad de Bartlett	544.02; $p = .000$	314.42; $p = .000$
Varianza Total Explicada	67.415	48.916 69.024

Finalmente, otra forma de evaluar la validez de un instrumento puede ser mediante un AFC. Como se aprecia en la Tabla 6, se obtuvieron mejores resultados para la opción de respuesta de concordancia a excepción de algunos índices de bondad de ajuste

poblacional y estadística, como lo son: χ^2/gl , la no significancia del modelo y RMSEA $< .08$ (Hair et al., 2019). Cabe señalar que estudios previos no suelen reportar χ^2/gl (e.g., Emerson et al., 2017).

Tabla 6. Índices de Bondad de Ajuste

	X ² /gl	p	RMR	GFI	NFI	IFI	RMSEA	TLI	CFI
Concordancia	3.847	.002	.55	.964	.965	.974	.116	.947	.974
Frecuencia	.481	.791	.051	.995	.992	.100	.000	.101	1.00

DISCUSIÓN

El objetivo fue comparar las opciones de respuesta tipo Likert de concordancia y de frecuencia. Para lograr lo anterior, se utilizó una escala que mide la satisfacción con la vida en trabajadores universitarios. A través del estudio de una muestra conformada por personal administrativo y docente, el presente estudio tuvo una connotación de novedad. Además, ésta puede incluir un gran potencial para aportar a la discusión sobre las características psicométricas de las escalas Likert (Marfeo et al., 2014). Los resultados obtenidos pueden servir de parteaguas para comprender mejor este tipo de escalas. Por ello, su discusión versa no precisamente sobre los análisis realizados con el SWLS precisamente, pero sí con otros estudios que comparan las opciones de respuesta tipo Likert.

En primer lugar, luego de haber contrastado los hallazgos de ambas opciones de respuesta -de concordancia y de frecuencia- por medio de la

comparación de medias, no se encontraron diferencias significativas, tal como lo mostró la prueba no paramétrica U de Mann-Whitney. Es entonces que, surge la interrogante de ¿Cuál opción de respuesta Likert es la mejor? En términos generales, la evidencia empírica no es clara de la existencia de diferencias significativas al comparar las opciones de respuesta. En cuanto a la literatura existente al respecto, se puede destacar a Brown (2004), quien, al estudiar las escalas de calificaciones con formatos de respuesta equilibrados, encontró pocas diferencias significativas al comparar características psicométricas de los ítems entre opciones de respuesta de concordancia y frecuencia; sin embargo, la primera mostró evidencia psicométrica más favorable para la medición de dicha variable (como en el presente estudio). Mientras que, para Tong et al. (2020), la opción de respuesta de frecuencia superó a la de concordancia al medir el burnout y compromiso; o en el caso de Mizumoto y Takeuchi (2009), quienes sí identificaron diferencias



estadísticamente significativas al comparar las medias de frecuencia y veracidad de la estrategia cognitiva y metacognitiva por medio de la prueba T, siendo la opción de respuesta de veracidad la que mostró mejores resultados. Esta técnica paramétrica busca mediante la comparación de medias (promedios) de dos muestras, comprobar que son estadísticamente independientes, siendo la U de Mann-Whitney su versión no paramétrica (Field, 2018). En cuanto a la comparación de la normalidad de las opciones de respuesta -mediante la simetría y curtosis- se encontraron niveles favorables, aspecto previamente estudiado por Brown (2004).

Cabe destacar que los hallazgos encontrados en las cargas factoriales de la opción de respuesta de concordancia en personal universitario obtuvo una agrupación adecuada, tomando como base su estructura unifactorial propuesta por Diener et al. (1985). Esto a pesar de que otras investigaciones habían detectado ciertos problemas con relación al ítem 5 en población argentina (Mikulic et al., 2019), lo cual no sucedió en este estudio. Sería recomendable continuar explorando la estructura factorial del SWLS, ya que las investigaciones recientes se han centrado en analizar la invarianza a través de métodos confirmatorios (e.g., Espejo et al., 2022), dejando de lado un aspecto relevante de la validez y la TRI, donde las escalas de medida son sensibles a los cambios de contexto (Diener et al., 2013).

Por otro lado, luego de haber llevado a cabo el análisis factorial, se encontró que la escala original propuesta por Diener et al. (1985), que emplea la opción de respuesta de concordancia, tuvo mejores resultados en cuanto al AFE y AFC. Poniendo énfasis en la validez del SWLS, los resultados de esta escala concuerdan con las revisiones de estudios empíricos a nivel global realizadas por Jang et al. (2017) y Emerson et al. (2017). Ahora bien, considerando estudios similares que hayan usado esta metodología para comparar escalas Likert -pero con otras variables- se destacan los resultados reportados por Mizumoto y Takeuchi (2009), quienes encontraron que la opción de respuesta de veracidad obtuvo mejores resultados en comparación con la de frecuencia, excepto a la no significancia de la X₂/gl. Se resalta que, con el fin de complementar los hallazgos sobre la validez del instrumento, se realizó también el análisis de validez convergente a través de la prueba de varianza media

extraída, de acuerdo con Hair et al. (2022), valores de $\geq .50$ son considerados aceptables, en este estudio solo la opción de respuesta de concordancia tuvo resultados satisfactorios ($AVE = .674$).

A partir de lo anterior, es importante que los investigadores tengan claridad de los supuestos que subyacen dentro de la teoría de medición (Jebb et al., 2021), como puede ser la TRI. Dicha teoría permite evaluar el desempeño de cada ítem, basándose en el supuesto de que está varía en relación con la capacidad de cada participante. Existiendo evidencia escasa al respecto, como es el caso de Moreta-Herrera et al. (2023) en población ecuatoriana y colombiana, y Nima et al. (2020) en población estadounidense, los autores concluyen que las propiedades psicométricas de las escalas variarán dependiendo del contexto situacional y contextual al que se enfrenten; en este sentido, la TRI permite considerar otras fuentes de error.

En cuanto a las cualidades psicométricas del SWLS, según Diener et al. (1985), ésta fue generada con el fin de hacer más precisa la medición de esta variable, pues otras escalas la habían intentado medir con un solo ítem. Si bien, es cierto que la mayor parte de la evidencia empírica de dicho instrumento ha seguido al pie de la letra la metodología propuesta por estos autores, han surgido problemas en cuanto a la invarianza de medida (Jovanović et al., 2022), la cual se centra en la validez de constructo del instrumento. Esto denota la importancia de tomar caminos alternativos para medir esta cualidad. Todo esto considerando que la construcción y/o adaptación de un instrumento de medición requiere poner atención en la teoría psicométrica y sus procedimientos (Jebb et al., 2021), siguiendo estos aspectos, se pretendió poner énfasis en la posibilidad de verificar la identidad de la versión original de concordancia de la escala; a partir de su comparación con la opción de respuesta de frecuencia, continuando algunas estrategias realizadas -en otras variables- en casos similares como lo realizado por Brown (2004), Marfeo et al. (2014), Mizumoto y Takeuchi (2009); los cuales son estudios pioneros metodológicamente hablando que pueden invitar a la reflexionar sobre la validez de constructo y usar caminos alternativos para verificar la validez de constructo de un instrumento de medición cuantitativo.



Considerando los estudios previos, con relación al uso de diferentes opciones de respuesta Likert en un mismo cuestionario, se identificó que este puede mostrar resultados diferentes (Mizumoto & Takeuchi, 2009); lo cual es un punto de partida para otras propuestas metodológicas como la de Marfeo et al. (2014), quienes recomiendan hacer una integración de escalas para mejorar la validez. No obstante, se resalta que la mejor opción de respuesta Likert es aquella que atiende al cumplimiento del objetivo de la investigación.

Finalmente, un aspecto a destacar de la investigación es la población de estudio –trabajadores universitarios – donde la mayoría de las investigaciones se han sesgado hacia los profesores (Emerson et al., 2017), si bien este estudio también exploró la perspectiva del personal académico, se incluyó al personal administrativo que ha sido rezagado. La relevancia del personal administrativo se debe a que estos se encuentran en constante contacto tanto con el personal académico, con los usuarios (estudiantes) y la comunidad, por las condiciones laborales es un trabajo demandante el cual puede influir en su nivel de satisfacción con la vida.

CONCLUSIONES

En la actualidad, el uso de la escala tipo Likert se ha popularizado (Anjaria, 2022); sin embargo, debido a su generalidad de uso, es posible que no se haga una reflexión profunda sobre las implicaciones de un incorrecto uso. Esto debido a que, si se llegara a emplear una opción de respuesta inadecuada, además de atentar contra la validez y confiabilidad del instrumento (supuestos psicométricos básicos), de manera práctica, se obtendrían resultados que no reflejaran la realidad del fenómeno de estudio. Finalmente, esto impactaría de manera negativa en

REFERENCIAS BIBLIOGRÁFICAS

1. Alimoradi, Z., Lin, C. Y., Ullah, I., Griffiths, M. D., & Pakpour, A. H. (2022). Item response theory analysis of the fear of COVID-19 scale (FCV-19S): A systematic review. *Psychology research and behavior management*, 15, 581-596. <https://doi.org/10.2147/PRBM.S350660>
2. 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>
3. Aryadoust, V., & Raquel, M. (2019). *Quantitative data analysis for language assessment volume II: advanced methods* (1st ed.). Routledge. <https://doi.org/10.4324/9781315187808>
4. Bertram, D. (2007). Likert scales. Retrieved November, 2(10), 1-10. https://www.researchgate.net/profile/Mahdi-Safarpour-2/post/what_is_a_logistic_regression_analysis/attachment/59d622fb79197b8077981515/A

cualquier tipo de estudio social donde se implemente este tipo de escala, ya sea a nivel individual, organizacional o social.

Las conclusiones a las que llegó el presente estudio se enumeran a continuación: (a) aunque no se encontraron diferencias estadísticamente significativas entre las opciones de respuesta de concordancia y de frecuencia del SWLS, los porcentajes de las respuestas se agruparon de manera similar a estudios previos, sin embargo, aunque ambas escalas mostraron niveles favorables casi en su totalidad en cuanto a asimetría y curtosis (normalidad), la escala de frecuencia obtuvo mejores resultados; (b) la versión de la escala Likert original del instrumento (de concordancia) obtuvo mejores resultados en cuanto a los análisis de validez de constructo y de confiabilidad; (c) la estrategia de comparación de escalas -como la del presente estudio- es algo que no había sido empleado para estudiar a un instrumento que cuenta con una vasta evidencia empírica sobre su validez, pero que todavía da pautas para nuevas metodologías alternativas; (d) la comparación de dos categorías de escala Likert, con el fin de medir la validez, puede ser una estrategia emergente que podría aportar a la discusión sobre los supuestos psicométricos para medir la satisfacción con la vida y otras variables, en especial en trabajadores universitarios; (e) la idoneidad de las opciones de la escala Likert se basa en el objetivo del instrumento de medición.

AGRADECIMIENTOS

Los autores agradecen el apoyo económico otorgado por parte del Instituto Tecnológico de Sonora a través del Programa de Fomento y Apoyo a Proyectos de Investigación (PROFAPI) con edición 2023.



- S%3A304626539139075%40144964003476
0/download/Likert+Scale+vs+Likert+Item.pdf
5. Bisquerra, R., & 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>
 6. Brown, G. T. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological reports*, 94(3), 1015-1024. <https://doi.org/10.2466/pr0.94.3.1015-1024>
 7. de Ayala, R. J. (2022). *The theory and practice of item response theory* (2^a ed.). The Guilford Press.
 8. 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
 9. Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social indicators research*, 112, 497-527. <https://doi.org/10.1007/s11205-012-0076-y>
 10. 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>
 11. Espejo, B., Martín-Carbonell, M., Checa, I., Paternina, Y., Fernández-Daza, M., Higuita, 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>
 12. Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (6^a ed.). SAGE Publications Ltd.
 13. González-Rico, P., Carvalho, V. S., Chambel, M. J., & Guerrero, E. (2018). Be well at work, be well outside work: A study with university workers. *Studies in Higher Education*, 43(6), 1034-1044. <https://doi.org/10.1080/03075079.2016.1212326>
 14. González-Rico, P., Guerrero-Barona, E., Chambel, M. J., & Guerrero-Molina, M. (2022). Well-being at work: Burnout and engagement profiles of university workers. *International Journal of Environmental Research and Public Health*, 19(23), 15436. <https://doi.org/10.3390/ijerph192315436>
 15. Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate Data Analysis* (7ma ed.). Cengage.
 16. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook* (6^a ed.). SAGE.
 17. 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)
 18. 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>
 19. 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., Ferreiro, 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>
 20. Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology*, 12, 637547. <https://doi.org/10.3389/fpsyg.2021.637547>
 21. 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>
 22. Jovanović, V., Rudnev, M., Arslan, G., Buzea,



- C., Dimitrova, R., Góngora, V., Guse, T., Ho, R. T. H., Iqbal, N., Jámbori, 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>
23. Marfeo, E. E., Ni, P., Chan, L., Rasch, E. K., & Jette, A. M. (2014). Combining agreement and frequency rating scales to optimize psychometrics in measuring behavioral health functioning. *Journal of clinical epidemiology*, 67(7), 781-784. <https://doi.org/10.1016/j.jclinepi.2014.02.005>
24. Mikulic, I. M., Crespi, M., & Caballero, R. Y. (2019). Escala de satisfacción con la vida (SWLS): Estudio de las propiedades psicométricas en adultos de Buenos Aires. *Anuario de Investigaciones*, 26, 395-402. <https://doi.org/10.17081/psico.22.42.3468>
25. Mizumoto, A., & Takeuchi, O. (2009). Comparing frequency and trueness scale descriptors in a Likert scale questionnaire on language learning strategies. *Journal Kiyo*, 12, 116-136. https://doi.org/10.20622/jltaj.12.0_116
26. Moreta-Herrera, R., Perdomo-Pérez, M., Reyes- Valenzuela, C., Gavilanes-Gómez, D., Rodas, J. A., & Rodríguez-Lorenzana, A. (2023). Analysis from the classical test theory and item response theory of the satisfaction with life scale (SWLS) in an Ecuadorian and Colombian sample. *Journal of Human Behavior in the Social Environment*, 1-13. <https://doi.org/10.1080/10911359.2023.2187915>
27. Nima, A. A., Cloninger, K. M., Persoon, B. N., Sikström, S., & García, D. (2020). Validation of subjective well-being measures using item response theory. *Frontiers in psychology*, 10, 3036. <https://doi.org/10.3389/fpsyg.2019.03036>
28. Roy, A. (2020). *A comprehensive guide for design, collection, analysis and presentation of Likert and other rating scale data: Analysis of Likert scale data*. Amazon Digital Services LLC.
29. South, L., Saffo, D., Vitek, O., Dunne, C., & Borkin, M. A. (2022). Effective use of Likert scales in visualization evaluations: A systematic review. *Computer Graphics Forum*, 41(3), 43-55. <https://doi.org/10.1111/cgf.14521>
30. Sürütü, L., Şeşen, H., & Maslakç, A. (2023). *Regression, mediation/moderation, and structural equation modeling with SPSS, AMOS, and PROCESS Macro*. Livre de Lyon.
31. Tong, J., Bickmeier, R. M., & Rogelberg, S. G. (2020). A comparison of frequency-and agreement-based response formats in the measurement of burnout and engagement. *International Journal of Environmental Research and Public Health*, 17(2), 543. <https://doi.org/10.3390/ijerph17020543>