



Revista Argentina de Ciencias del Comportamiento
ISSN: 1852-4206
paulaabate@gmail.com
Universidad Nacional de Córdoba
Argentina

Calleja, Nazira; Candelario Mosco, Julieta Beatriz;
Rosas Medina, Jorge Humberto; Souza Colín, Erika
Equivalencia psicométrica de las aplicaciones impresas y electrónicas de tres escalas psicosociales
Revista Argentina de Ciencias del Comportamiento,
vol. 12, núm. 2, 2020, Mayo-Agosto, pp. 50-58
Universidad Nacional de Córdoba
Córdoba, Argentina

Disponible en: <https://www.redalyc.org/articulo.oa?id=333468469001>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Equivalencia psicométrica de las aplicaciones impresas y electrónicas de tres escalas psicosociales

Calleja, Nazira ^{*, a}; Candelario Mosco, Julieta Beatriz^a; Rosas Medina, Jorge Humberto^a; Souza Colín, Erika^a

Artículo Metodológico

Resumen	Abstract	Tabla de Contenido
<p>El propósito del estudio fue examinar el efecto del formato de aplicación (impreso versus electrónico) de escalas psicosociales sobre sus propiedades psicométricas y las mediciones obtenidas. Estudiantes universitarios ($N = 435$) respondieron tres escalas breves y unidimensionales, que evaluaban optimismo, gratitud y bienestar subjetivo, en formato impreso (52%) o electrónico (48%). Se encontró que las medias de los puntajes de las tres escalas no difirieron significativamente al comparar ambos medios de aplicación. En general, tanto los instrumentos electrónicos como los de papel aportaron resultados con propiedades similares de confiabilidad y validez. Al analizar la invarianza de medición, comparando el modelo sin restricciones en la estructura factorial con los distintos modelos con invarianza, se comprobó que las medidas eran equivalentes en ambos formatos. Se concluyó que el medio de aplicación no afecta los índices obtenidos, por lo que podrían emplearse indistintamente.</p>	<p>Psychometric equivalence of printed and electronic administration of three psychosocial scales. The purpose of the study was to examine the influence of the format of administration (printed versus electronic) of psychosocial scales on their psychometric properties and the measurements obtained. University students ($N = 435$) completed three short and one-dimensional scales, which evaluated optimism, gratitude, and subjective well-being, in printed (52%) or electronic (48%) format. It was found that the mean scores of the three scales did not differ significantly when comparing both modes of application; overall, both electronic and paper instruments provided results with similar properties of reliability and validity. When analyzing the measurement invariance, comparing the model without constraints in the factor structure with the different models with invariance, the measures were equivalent in printed and electronic formats. It was concluded that the format of administration does not affect the indices obtained, so they could be used interchangeably.</p>	<p>Introducción 50 Método 51 Participantes 51 Diseño 51 Instrumentos 52 Procedimiento 52 Análisis de los datos 52 Consideraciones éticas 53 Resultados 53 Discusión 56 Referencias 57</p>
<p>Palabras clave: aplicaciones electrónicas, escalas Likert, validez, invarianza de medición.</p>	<p>Keywords: electronic administration, Likert scales, validity, measurement invariance.</p>	

Recibido el 28 agosto de 2019; Aceptado el 30 de diciembre de 2019

Editaron este artículo: Raquel Peltzer, Paula Abate, Sebastián Miranda y María Victoria Ortiz

Introducción

En la actualidad, muchos investigadores están recurriendo cada vez con mayor frecuencia a las Tecnologías de la Información y la Comunicación (TIC) para efectuar la aplicación de sus instrumentos, con el propósito de ampliar la cantidad de participantes a los que se tiene acceso y reducir costos. En este contexto, diversos autores se han abocado a estudiar la variación de las propiedades psicométricas de las

escalas, dependiendo de si su medio de aplicación es electrónico o de papel y lápiz (Jeong, 2014; Noyes & Garland, 2008).

Muehlhausen et al. (2015) efectuaron una revisión sistemática y un meta-análisis de estudios del área médica sobre la equivalencia entre aplicaciones impresas y en aparatos electrónicos (tabletas, teléfonos celulares y laptops) de cuestionarios para evaluar síntomas clínicos (v.gr.,

^a Universidad Nacional Autónoma de México, Facultad de Psicología. México.

*Enviar correspondencia a: Calleja, N. E-mail: ncalleja@unam.mx

dolor crónico, asma, disfunción sexual, enfermedad de Parkinson, funcionamiento pulmonar, osteoartritis); sus resultados mostraron que ambos formatos eran cuantitativamente comparables. Cuando [Determann, Lambooi, Steyeberg, Bekker-Grob y Ardine de Witt \(2017\)](#) aplicaron encuestas sobre seguros de salud, el modo de administración tampoco tuvo efectos sobre la calidad de los datos. Hallazgos similares se han reportado en el área educativa, en pruebas de conocimientos. [Hosseini, Abidin y Baghdarnia \(2014\)](#) aplicaron pruebas de comprensión lectora en inglés, en tanto que [Karay, Schaubert, Stoch y Schüttelz-Brauns \(2015\)](#) evaluaron el desempeño de estudiantes de medicina; en ninguno de estos estudios se obtuvieron diferencias significativas entre las aplicaciones en papel y en línea, por lo que los autores concluyeron que la aplicación por computadora es un modo viable de realizar evaluaciones de desempeño académico, además de ser preferido por los participantes.

En cuanto a la medición de constructos psicológicos, [Read, Farrow, Jaanimägi y Ouimette \(2009\)](#) evaluaron, tanto en papel y lápiz como en versión electrónica, el trastorno de estrés postraumático, y [Bagby, Ayearst, Morariu, Watters y Taylor \(2013\)](#) midieron alexitimia en las dos versiones; en ambos estudios los resultados mostraron que los instrumentos de papel y los electrónicos podían utilizarse indistintamente, ya que ambos correlacionaron fuertemente y presentaron índices de confiabilidad y validez adecuados.

Sin embargo, otras investigaciones han reportado hallazgos contrarios. Por ejemplo, [Jeong \(2014\)](#), al aplicar diversas pruebas de conocimientos a estudiantes coreanos, encontró que en los cuestionarios de papel se obtuvieron puntajes más altos en comparación con los electrónicos; el autor explica que las diferencias podrían haberse debido al modo en el que se mostró el contenido en la pantalla. [Noyes y Garland \(2008\)](#) realizaron una revisión de la literatura que compara las aplicaciones de papel contra las electrónicas y concluyeron que, si bien no siempre es posible lograr la equivalencia entre ellas, el avance en la tecnología ha llevado a lograr mayor semejanza.

Una ventaja indiscutible de las aplicaciones efectuadas a través de medios electrónicos tiene que ver con los datos perdidos. La pérdida o no disponibilidad de una parte de los datos de los participantes es un problema común en la

investigación ([Cuesta, Fonseca-Pedrero, Vallejo, & Muñiz, 2013](#)). Los datos perdidos, según sus características y proporción, pueden afectar las estimaciones de la medida del constructo en cuestión, tanto por la reducción del número de casos o de variables, como por los posibles sesgos debidos a la pérdida de valores en participantes con características específicas ([Dong & Peng, 2013](#); [Engels & Diehr, 2003](#)). Para minimizarlos, los investigadores implementan estrategias, como aumentar el tamaño de la muestra, monitorear la recolección de los datos, valorar la proporción y características de los datos perdidos y efectuar con cautela procedimientos de imputación ([Durán, 2005](#)). Sin embargo, como señala [Allison \(2002\)](#), la mejor solución al problema de los datos perdidos es no tenerlos, y esto es factible cuando se hacen aplicaciones electrónicas, al impedir el avance al siguiente reactivo hasta no haber contestado el anterior.

[Streiner, Norman y Cairney \(2015\)](#) señalan que otra ventaja de utilizar medios electrónicos para las aplicaciones es permitir a los participantes expresarse con mayor confianza. Al respecto, [Spark et al. \(2015\)](#), en su estudio sobre comportamiento sexual, obtuvieron un mayor número de preguntas respondidas en la versión electrónica que en la de papel. [Determann et al. \(2017\)](#) y [Richardson, Johnson, Ratner y Zumbo \(2009\)](#) señalan, además, que las aplicaciones en línea requieren un menor tiempo de respuesta que las de papel y lápiz.

El objetivo del presente estudio fue comparar, en población mexicana, las aplicaciones efectuadas vía electrónica versus papel, no sólo de una sino de tres escalas breves que evalúan constructos psicosociales (optimismo, gratitud y bienestar psicológico), respecto de varios indicadores: medidas descriptivas e índices de confiabilidad, así como evidencias de validez vinculadas a la estructura interna e invarianza de medición, la cual ha sido escasamente reportada en este tipo de estudios.

Método

Participantes

Participaron en el estudio 435 estudiantes universitarios de licenciatura (82.1%) y posgrado (17.9%) de la Ciudad de México, con una media de edad de 22.30 ± 3.33 años (rango de 18 a 30 años); 48.7% hombres y 51.3% mujeres.

Diseño

Se empleó un diseño de comparación de dos grupos no aleatorios; 52.0% de los participantes formaron parte del grupo de aplicación impresa y 48.0%, del de aplicación electrónica. No hubo diferencias estadísticamente significativas en la edad de los participantes de ambos grupos, $t(433) = 1.574$, $p = .116$, ni en la proporción de hombres y mujeres, $\chi^2(1) = .068$, $p = .794$.

Instrumentos

Cuestionario de Optimismo, de Pedrosa, Celis-Atenas, Suárez-Álvarez, García-Cueto y Muñiz (2015). Mide la tendencia personal estable de evaluar positivamente los posibles eventos futuros. Su único factor con nueve reactivos explica el 42.43% de la varianza total, con un índice de consistencia interna α de Cronbach de .84.

Cuestionario de Gratitud, de McCullough, Emmons y Tsang (2002). Evalúa el reconocimiento de los actos bondadosos de las personas y la manifestación de agradecimiento. Está conformado por seis reactivos integrados en un factor. El análisis factorial confirmatorio probó su ajuste unidimensional (CFI = .95, SRMR = .04); se reportó un coeficiente alfa de .82. La versión en español quedó integrada por siete reactivos, ya que uno de ellos se dividió en virtud de que contenía dos ideas diferentes.

Escala de Bienestar Subjetivo (EBS-8), de Calleja y Mason (2020). Mide el bienestar subjetivo, definido como la satisfacción con la propia vida y la presencia de estados afectivos positivos. Consta de un factor con ocho reactivos, que explica el 78.73% de la varianza. Los índices de ajuste del análisis factorial confirmatorio resultaron satisfactorios (GFI = .980, NFI = .992; CFI = .998; RMSEA = .041). La confiabilidad de las puntuaciones fue estimada mediante el método de consistencia interna, observándose un coeficiente α de Cronbach de .971.

Procedimiento

Los reactivos de las tres escalas fueron integrados de forma aleatoria en una batería, que se elaboró tanto en formato impreso como en versión electrónica. Las opciones de respuesta fueron: *Muy en desacuerdo* (0), *En desacuerdo* (1), *Ni en acuerdo ni en desacuerdo* (2), *De acuerdo* (3), *Muy de acuerdo* (4), *Bastante de acuerdo* (5) y *Totalmente de acuerdo* (6). Se realizaron pruebas piloto de ambas versiones con

participantes cuyas características eran similares a los de los integrantes de la muestra; se efectuó un laboratorio cognitivo para cada versión. Esta metodología cualitativa, que permite estudiar el proceso mental que las personas utilizan cuando resuelven una tarea, ha sido empleada exitosamente en el desarrollo y adaptación de instrumentos de medición (Zucker, Sassman, & Case, 2004). En los laboratorios cognitivos se exploró la comprensión, interpretación y pertinencia de cada uno de los reactivos y de las opciones de respuesta, así como la estructura de la batería. Se detectaron errores en la redacción de tres de los 24 reactivos; la secuencia y el formato de respuesta no presentaron problemas. Los constructos medidos resultaron de interés y fueron pertinentes para los integrantes de la muestra.

Los instrumentos impresos fueron aplicados a estudiantes universitarios en sus propios salones de clase o en el campus de la institución a la que asistían; en el primer caso, en forma grupal y en el segundo, de manera individual. Después de presentarse, los aplicadores explicaban los objetivos del estudio y los invitaba a participar en él de manera voluntaria; se hacía énfasis en que sus respuestas serían tratadas de forma anónima y confidencial; finalmente, se solicitaba su consentimiento informado. Las aplicaciones electrónicas fueron respondidas en las plataformas Google Forms y E-Survey Creator, y en los formularios de Adobe Acrobat IX. La difusión se efectuó a través de las redes sociales Facebook y Twitter de los autores, o por invitación directa vía correo electrónico a sus contactos. Se explicaba brevemente el propósito de la investigación y los criterios de inclusión de la muestra: estudiantes universitarios de licenciatura o posgrado, hombres y mujeres, de 18 a 30 años de edad. Se les solicitaba participar en el estudio y difundirlo entre sus contactos; se señalaba su carácter completamente anónimo y voluntario; finalmente, se incluía el link de la batería.

Análisis de datos

Se analizaron las propiedades psicométricas de los tres instrumentos por medio de aplicación (impreso y electrónico). Se efectuaron los siguientes análisis estadísticos: frecuencia de valores perdidos, medias y desviaciones estándar y pruebas t para muestras independientes. La confiabilidad de los puntajes de los instrumentos

se obtuvo mediante índices de consistencia interna (α de Cronbach) y de confiabilidad ordinal o compuesta ω , para la que Raykov y Shrout (2002) sugieren un índice mínimo de 0.70. Con el propósito de probar la igualdad de los coeficientes alfa obtenidos para los puntajes de los instrumentos en el formato impreso y en el electrónico, se aplicó el estadístico W de Feldt (1969) para dos muestras independientes. Para confirmar la estructura factorial y determinar el porcentaje de varianza media extraída (AVE) por el factor, se efectuaron análisis factoriales confirmatorios (AFC) con el método de máxima verosimilitud (Schermelleh-Engel, Moosbrugger, & Müller, 2003); para determinar el ajuste del modelo propuesto, se analizaron los siguientes índices (Hu & Bentler, 1999): prueba de bondad de ajuste χ^2 ; cociente χ^2/df (CMIN/DF) cuyo valor <3 indica un ajuste adecuado, así como el índice comparativo CFI y el de error RMSEA, que muestran un buen ajuste si su valor es $> .95$ y $< .08$, respectivamente. Se calculó la varianza media extraída a partir de las λ obtenidas en análisis factoriales confirmatorios, cuyo valor, recomendado por Fornell y Larcker (1981), debe estar por encima de .50. Se efectuaron AFC multi-grupo para probar la invarianza de medición por medios de aplicación (Vandenberg & Lance, 2000). Se probaron cuatro tipos de invarianza: de configuración (no variabilidad de la estructura factorial entre los grupos, mientras se calculan libremente las cargas factoriales, los interceptos y los residuales), métrica (restricción de la

estructura factorial y las cargas factoriales), escalar (restricción, además de los dos anteriores, de los interceptos) y estricta (restricción de los residuales, además de los tres índices anteriores). Para evaluar las diferencias entre los modelos, se utilizaron como índices los cambios en CFI y en RMSEA; se apoya una fuerte invarianza cuando el ΔCFI es ≤ 0.01 , el $\Delta RMSEA$ es ≤ 0.015 y el $\Delta \chi^2$ resulta con $p > .05$ (Cheung & Rensvold, 2002). Se utilizaron los programas SPSS y AMOS, ambos versión 22, para llevar a cabo los procedimientos estadísticos.

Consideraciones éticas

Este trabajo forma parte de la línea de investigación “Mejores prácticas en la medición de constructos psicosociales y de la salud”, de la Facultad de Psicología de la Universidad Nacional Autónoma de México. El estudio se ha realizado considerando los principios éticos de la Declaración de Helsinki (World Medical Association, 2013).

Resultados

Medias y desviaciones estándar

No se encontraron diferencias estadísticamente significativas entre las medias de la aplicación impresa y la electrónica en ninguna de las escalas aplicadas. Los valores de la d de Cohen y del tamaño del efecto (r) resultaron mínimos, indicando que el medio de aplicación prácticamente no tiene efecto sobre los puntajes obtenidos en las escalas (tabla 1).

Tabla 1.

Media (desviación estándar) para los dos medios de aplicación, prueba estadística de las diferencias entre ellos, d de Cohen y tamaño del efecto.

Instrumento	Medio de aplicación		Prueba t	d de Cohen	Tamaño del efecto r
	Impreso	Electrónico			
Optimismo	4.08 (1.19)	4.18 (1.16)	$t_{(433)} = .898$, $p = .369$.086	.043
Gratitud	4.41 (1.13)	4.54 (1.12)	$t_{(433)} = 1.167$, $p = .244$.116	.057
Bienestar subjetivo	4.27 (1.27)	4.23 (1.17)	$t_{(433)} = 0.353$ $p = .724$.032	.016

Nota. Media teórica = 3; rango = 1 a 7.

Confiabilidad

La tabla 2 muestra los índices de consistencia interna total (α de Cronbach) y los de confiabilidad ordinal o compuesta obtenidos para los instrumentos

aplicados en formato impreso y electrónico. Para las tres escalas, los índices fueron altos; los valores de la confiabilidad compuesta ω estuvieron por encima del límite mínimo establecido de .70.

Las diferencias de los índices alfa entre ambos formatos, al probarlas con la prueba *W* de Feldt, resultaron estadísticamente no significativas.

Tabla 2.

Índices de consistencia interna (α de Cronbach) y confiabilidad ordinal o compuesta ω de los dos medios de aplicación para cada uno de los instrumentos; prueba de diferencias entre los coeficientes alfa y significancia.

Instrumento	α de Cronbach			Confiabilidad ordinal o compuesta ω	
	Impreso (<i>n</i> = 226)	Electrónico (<i>n</i> = 209)	<i>W</i> de Feldt	Impreso	Electrónico
Optimismo	.934	.926	<i>W</i> = .891, <i>p</i> = .802	.945	.939
Gratitud	.891	.912	<i>W</i> = .807, <i>p</i> = .943	.955	.938
Bienestar subjetivo	.940	.943	<i>W</i> = .950, <i>p</i> = .647	.954	.947

Varianza media extraída (AVE)

Los AFC ratificaron las soluciones unifactoriales para las tres escalas, con índices de ajuste por arriba de los criterios de corte (descritos en la sección de Análisis de los datos), los cuales se muestran en los modelos de línea base de la tabla 3. En todos los casos, la varianza media extraída resultó $> .60$ y fue similar en las aplicaciones electrónicas e impresas; en las escalas de Optimismo y Bienestar subjetivo fueron ligeramente más altas en las versiones impresas (.631 y .726, respectivamente) que en las electrónicas (.620 y .690, respectivamente), en tanto que, en la escala de Gratitud, el valor fue levemente más alto en el formato electrónico (.632) que en el impreso (.607).

Invarianza de medición

Se efectuaron AFC multi-grupo para probar los modelos de invarianza (tabla 3). Inicialmente, se probó la invarianza de configuración, línea base o libre (M1), que proponía que las escalas tendrían una estructura unifactorial en ambos medios de aplicación y se permitió que las cargas factoriales, los interceptos y las varianzas de error se estimaran libremente. Los resultados apoyaron el buen ajuste de los reactivos de cada escala a la unidimensionalidad. A continuación, se probaron los modelos de invarianza métrica (M2), en los que se restringieron las cargas factoriales para

que fueran iguales entre la aplicación impresa y la electrónica. Los índices mostraron que los modelos ajustaron bien y, cuando se compararon con los del M1, los incrementos fueron menores a los criterios establecidos: el ΔCFI fue ≤ 0.01 , el $\Delta RMSEA$ resultó ≤ 0.015 , y $\Delta \chi^2$ fue no significativo ($p < .05$). La prueba de los modelos de invarianza escalar (M3), en los que los interceptos, además de las cargas factoriales, se restringieron para que fueran iguales entre ambas aplicaciones, mostraron un buen ajuste. Al compararlos con los del M2, no se presentaron cambios significativos en CFI y RMSEA, ni en χ^2 . Finalmente, los modelos de invarianza estricta (M4), en los que se restringieron, además de las cargas factoriales y los interceptos, las varianzas de error, también ajustaron correctamente y, comparados con los del M3, los incrementos en los índices no resultaron significativos, excepto en $\Delta \chi^2$ de la Escala de Bienestar subjetivo [$\Delta \chi^2 = 33.412$ (15), $p = .004$]. En conjunto, los datos indicaron que cuando los elementos de la estructura factorial se mantuvieron invariantes en función de los medios de aplicación, los índices de ajuste fueron predominantemente comparables, por lo que la medición del optimismo, la gratitud y el bienestar subjetivo no varía cuando los instrumentos se aplican en formato impreso y digital. Por tanto, los puntajes son comparables entre ambos medios de aplicación.

Tabla 3.

Índices de ajuste para la prueba de los modelos de invarianza de las escalas de Optimismo, Gratitud y Bienestar subjetivo, por medio de aplicación (impreso o electrónico)

Modelo	$\chi^2(gl)$	χ^2/gl	CFI	RMSEA [IC 90%]	Compa- ración	$\Delta\chi^2$	ΔCFI	$\Delta RMSEA$
<i>Optimismo</i>								
M1. Invarianza de configuración (Línea base)	93.709 (42)	2.231	.981	.052 [.038-.067]				
M2. Invarianza métrica o débil (λ restringidas)	108.701 (50)	2.174	.978	.052 [.039-.065]	M2 vs M1	14.992 (8), $p = .059$	-.003	.000
M3. Invarianza escalar o fuerte (λ y τ restringidos)	119.586 (59)	2.027	.977	.049 [.036-.061]	M3 vs M2	10.885 (9), $p = .284$	-.001	-.002
M4. Invarianza estricta (λ , τ y θ restringidos)	139.374 (33)	1.858	.976	.045 [.033-.056]	M4 Vs M3	19.788 (26), $p = .801$	-.001	-.004
<i>Gratitud</i>								
M1. Invarianza de configuración (Línea base)	33.205 (16)	2.075	.990	.050 [.025-.074]				
M2. Invarianza métrica o débil (λ restringidas)	36.219 (22)	1.646	.992	.039 [.013-.060]	M2 vs M1	3.014 (6), $p = .807$.002	-.011
M3. Invarianza escalar o fuerte (λ y τ restringidos)	40.751 (29)	1.405	.993	.031 [.000-.051]	M3 vs M2	4.532 (7), $p = .717$.001	-.008
M4. Invarianza estricta (λ , τ y θ restringidos)	63.342 (43)	1.473	.986	.035 [.017-.052]	M4 Vs M3	22.591 (14), $p = .067$	-.007	-.004
<i>Bienestar subjetivo</i>								
M1. Invarianza de configuración (Línea base)	48.186 (28)	1.721	.993	.041 [.020-.060]				
M2. Invarianza métrica o débil (λ restringidas)	53.998 (35)	1.543	.994	.053 [.014-.042]	M2 vs M1	5.812 (7), $p = .562$.001	.007
M3. Invarianza escalar o fuerte (λ y τ restringidos)	59.444 (43)	1.382	.994	.047 [.002-.041]	M3 vs M2	5.446 (8), $p = .709$.000	-.006
M4. Invarianza estricta (λ , τ y θ restringidos)	92.856 (58)	1.601	.988	.051 [.022-.041]	M4 Vs M3	33.412 (15), $p = .004$.006	.004

Nota. λ = Cargas factoriales; τ = Interceptos; θ = Varianzas de error; CFI = Índice de Ajuste Comparativo; RMSEA = Raíz del Error Cuadrático Medio.

Valores perdidos

En las aplicaciones de papel y lápiz, el porcentaje de valores perdidos fue de 5.16, considerando la batería completa de 24 reactivos. El porcentaje de no respuestas fue mayor en las escalas de Gratitud (1.27%, con 7 reactivos) y

Bienestar subjetivo (1.25%, con 8 reactivos) que en la de Optimismo (0.63%, con 9 reactivos). En las aplicaciones electrónicas no se presentaron valores perdidos, puesto que se requería contestar cada reactivo para poder continuar con el siguiente.

Discusión

La finalidad del presente estudio fue determinar el efecto del medio de aplicación (impresa o electrónica) de escalas tipo Likert sobre la calidad psicométrica y los resultados de la medición de tres constructos psicológicos: optimismo, gratitud y bienestar subjetivo. Los criterios de evaluación fueron: estadísticos descriptivos (media y desviación estándar), confiabilidad (α de Cronbach y confiabilidad ordinal o compuesta), varianza media extraída (evidencias de validez vinculadas a la estructura interna) e invarianza de medición.

Se esperaba que las medias obtenidas fueran similares, independientemente del medio de aplicación que se utilizara. Los resultados no mostraron diferencias significativas entre las medias de la aplicación electrónica y la impresa de las tres escalas (optimismo, gratitud y bienestar subjetivo). Estos hallazgos concuerdan con los obtenidos por Bagby et al. (2013), Read et al. (2009) y Karay et al. (2015), en cuyos estudios los puntajes de las aplicaciones en línea y en papel tampoco difirieron.

Tanto en las aplicaciones electrónicas como en las impresas, los índices de confiabilidad obtenidos (α de Cronbach y confiabilidad compuesta ω) fueron altos, y las diferencias entre los coeficientes alfa, no significativas. Por tanto, es probable que ambos medios de aplicación resulten equivalentes en confiabilidad, similitud que ha sido reportada, entre otros autores, por Bagby et al. (2013) y Read et al. (2009).

Respecto de las evidencias de validez vinculadas a la estructura interna, se encontró la misma estructura unifactorial en ambos medios de aplicación. La varianza media extraída fue similarmente alta, con diferencias mínimas entre ambos formatos. Otros estudios de los últimos años han indicado también esta equivalencia (cfr., Determann et al., 2017; Karay et al., 2015).

Al analizar la invarianza factorial de las tres escalas para evaluar el grado de igualdad de su estructura entre las aplicaciones impresas y electrónicas, se encontró que los cuatro modelos puestos a prueba eran básicamente comparables cuando los elementos sucesivos de la conformación factorial (cargas factoriales, interceptos y residuales) se mantenían invariantes a través del medio de aplicación, con la única excepción del $\Delta\chi^2$ en la invarianza estricta de la escala de Bienestar

subjetivo. En este caso sería asumible una invarianza parcial (Dimitrov, 2010), ya que el resto de los índices muestran invarianza métrica; por otra parte, se ha reconocido que las pruebas de invarianza estricta son excesivamente restrictivas (Bentler, 2004). Por tanto, con base en los resultados obtenidos, es posible comparar las puntuaciones medias de las escalas en ambas modalidades de aplicación. Entre los pocos estudios que han probado la invarianza de medición entre medios de aplicación, el de Richardson et al. (2009) llegó a una conclusión similar; los autores, al examinar las respuestas de adolescentes a la Escala de Dependencia al Tabaco en versiones en papel y en la web, encontraron que las pruebas de invarianza de medición indicaron que el modo de administración no influyó en el funcionamiento psicométrico de la escala.

Los datos perdidos en las aplicaciones impresas fueron de alrededor del cinco por ciento. Sin embargo, en las aplicaciones en línea no era posible pasar al siguiente reactivo hasta no responder el anterior, por lo que el número de valores perdidos fue nulo, característica muy ventajosa de las aplicaciones electrónicas, dados los problemas que implican estos datos en la investigación: reducción del número de casos o de variables, sesgos en los datos, así como la necesidad de evaluar sus características y de hacer imputaciones estadísticas. Además, se consiguen resultados con rapidez, se evitan los errores de captura y se reduce los costos (Dong & Peng, 2013; Durán, 2005). A pesar de estas ventajas, habrá que considerar que un aspecto problemático de las aplicaciones en línea es la variabilidad en la tasa de respuesta, que puede oscilar entre el 9% y el 94% (Braithwaite, Emery, De Lusignan, & Sutton, 2003) y que puede implicar sesgos debido a su posible asociación con variables como las características de los participantes, el diseño del estudio, el interés del tema y la habilidad en el uso de Internet por el encuestado (Aerny Perreten et al., 2012).

En suma, las diversas evidencias obtenidas en el presente estudio sugieren que las aplicaciones electrónicas de escalas psicosociales poseen propiedades psicométricas similares a las de papel y que, por lo tanto, podrían utilizarse intercambiabilmente.

Limitaciones y propuestas

Los resultados de este estudio muestran que tanto los instrumentos electrónicos como los de papel pueden aportar resultados igualmente confiables y válidos; no obstante, la aplicación electrónica evita los datos perdidos, situación muy ventajosa en la investigación.

Una limitación del presente estudio es que se realizó sólo con estudiantes universitarios, cuyo uso frecuente de las TIC pudo haber influido en los resultados obtenidos, por lo que será necesario efectuar estudios semejantes con muestras de población abierta, de diferentes edades, niveles educativos y ocupaciones. Puesto que aquí se aplicaron escalas unidimensionales, cortas, de sólo tres constructos, se sugiere llevar a cabo estudios con instrumentos más largos, multidimensionales y que evalúen variables diversas. Así también, sería necesario preguntar a los participantes su preferencia respecto de los dos medios de aplicación y la posibilidad de que se expresen con confianza, así como evaluar los tiempos de respuesta. En el presente estudio no se evaluó la tasa de respuesta; en virtud de que la baja participación, en particular en las aplicaciones electrónicas, puede implicar sesgos y disminuir la posibilidad de generalizar los resultados, se sugiere medirla en futuros estudios y determinar su posible asociación con otras variables. Finalmente, habrá que considerar que los elementos psicométricos estudiados aquí son solo un aspecto del extraordinariamente complejo proceso de la medición del comportamiento humano.

Referencias

- Aerny Perreten, N., Domínguez-Berjón, M. F., Astray Mochales, J., Esteban-Vasallo, M. D., Blanco Ancos, L. M., & Lópaz Pérez, M. Á. (2012). Tasas de respuesta a tres estudios de opinión realizados mediante cuestionarios en línea en el ámbito sanitario. *Gaceta Sanitaria*, 26(5), 477-479. doi: 10.1016/j.gaceta.2011.10.016
- Allison, P. D. (2002). *Missing Data*. Thousand Oaks, CA: SAGE Publications.
- Bagby, M., Ayearst, L., Morariu, R., Watters, C., & Taylor, G. (2013). The Internet administration version of the 20-Item Toronto Alexithymia scale. *Psychological Assessment*, 26(1), 16-22. doi: 10.1037/a0034316
- Bentler, P. M. (2004). *EQS 6: Structural equation program manual*. Encino, CA: Multivariate Software.
- Braithwaite, D., Emery, J., De Lusignan, S., & Sutton, S. (2003). Using the Internet to conduct surveys of health professionals: A valid alternative? *Family Practice*, 20(5), 545-551. doi: 10.1093/fampra/cmg509
- Calleja, N., & Mason, T. (2020). Escala de Bienestar Subjetivo (EBS-20 y EBS-8): Construcción, validación e invarianza de medición. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica –e Avaliação Psicológica, RIDEP*, 55(2), 185-201. doi: 10.21865/RIDEP55.2.14
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. doi: 10.1097/NNR.0b013e3182544750
- Cuesta, M., Fonseca-Pedrero, E., Vallejo, G., & Muñiz, J. (2013). Datos perdidos y propiedades psicométricas en los tests de personalidad. *Anales de Psicología*, 29(1), 285-292. doi: 10.6018/analesps.29.1.137901
- Determann, D., Lambooi, M., Steyerberg, E., Bekker-Grob, E., & Ardine de Wit, G. (2017). Impact of survey administration mode on the results of a health-related discrete choice experiment: Online and paper comparison. *Value in Health*, 20(7), 953-960. doi: 10.1016/j.jval.2017.02.007
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121-149. doi: 10.1177/0748175610373459
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2, 222. doi: 10.1186/2193-1801-2-222
- Durán, P. (2005). Los datos perdidos en estudios de investigación, ¿son realmente datos perdidos? *Archivos Argentinos de Pediatría*, 103(6), 566-568. Recuperado de: <https://www.sap.org.ar/docs/publicaciones/archivosarg/2005/566.pdf>
- Engels, J. M., & Diehr, P. (2003). Imputation of missing longitudinal data: A comparison of methods. *Journal of Clinical Epidemiology*, 56(10), 968-976. doi: 10.1016/S0895-4356(03)00170-7
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*, 34(3), 363-373. doi: 10.1007/BF02289364
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. doi: 10.1177/002224378101800104
- Hosseini, M., Abidin, M., & Baghdarnia, M. (2014). Comparability of test results of computer based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran. *Procedia - Social and Behavioral Sciences*, 98, 659-667. doi: 10.1016/j.sbspro.2014.03.465
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit

- indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi: 10.1080/10705519909540118
- Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behavior & Information Technology*, 33(4), 410-422. doi: 10.1080/0144929X.2012.710647
- Karay, Y., Schaubert, S., Stoch, C., & Schüttpeitz-Brauns, K. (2015). Computer versus paper - does it make any difference in test performance? *Teaching and Learning in Medicine*, 27(1), 57-62. doi: 10.1080/10401334.2014.979175
- McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, 82(1), 112-127. doi: 10.1037//0022-3514.82.1.112
- Muehlhausen, W., Doll, H., Quadri, N., Fordham, B., O'Donohoe, P., Dogar, N., & Wild, D. J. (2015). Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health and Quality of Life Outcomes*, 13(1), 167. doi: 10.1186/s12955-015-0362-x
- Noyes, J., & Garland, K. (2008). Computer vs. paper based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352-1375. doi: 10.1080/00140130802170387
- Pedrosa, I., Celis-Atenas, K., Suárez-Álvarez, J., García-Cueto, E., & Muñiz, J. (2015). Cuestionario para la evaluación del optimismo: Fiabilidad y evidencias de validez. *Terapia psicológica*, 33(2), 127-138. doi: 10.4067/S0718-48082015000200007
- Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 195-212. doi: 10.1207/S15328007SEM0902_3
- Read, J., Farrow, S., Jaanimägi, F., & Ouimette, P. (2009). Assessing trauma and traumatic stress via the Internet: Measurement equivalence and participant reactions. *Traumatology*, 15(1), 94-102. doi: 10.1177/1534765608325121
- Richardson, C. G., Johnson, J. L., Ratner, P. A., & Zumbo, B. D. (2009). The influence of web-versus paper-based formats on the assessment of tobacco dependence: Evaluating the measurement invariance of the Dimensions of Tobacco Dependence Scale. *Substance Abuse: Research and Treatment*, 3, 1-14. doi: 10.4137/SART.S960
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of Structural Equation Models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74. doi: 10.1207/S15328007SEM0902_3
- Spark, S., Lewis, D., Vaisey, A., Smyth, E., Wood, A., Temple-Smith, M. ... Hocking, J. (2015). Using computer-assisted survey instruments instead of paper and pencil increased completeness of self-administered sexual behavior questionnaires. *Journal of Clinical Epidemiology*, 68(1), 94-101. doi: 10.1016/j.jclinepi.2014.09.011
- Streiner, D., Norman, G., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5^a ed.). Oxford: Oxford University Press.
- Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70. doi: 10.1177/109442810031002
- World Medical Association (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *American Journal of Medicine*, 310(20), 2191-2194. doi: 10.1001/jama.2013.281053
- Zucker, S., Sassman, C., & Case, B. J. (2004). *Cognitive labs*. San Antonio, TX: Hartcourt Assessment Inc.