

# ESTRUCTURA FACTORIAL DEL CUESTIONARIO DE SALUD GENERAL GHQ-12 EN POBLACIÓN GENERAL DE MÉXICO

## FACTOR STRUCTURE OF THE GHQ-12 GENERAL HEALTH QUESTIONNAIRE IN MEXICAN POPULATION

DOI: 10.22199/S07187475.2016.0001.00004

Recibido: 07 de marzo de 2016 | Aceptado: 31 de marzo de 2016

PEDRO SOLÍS-CÁMARA <sup>1</sup>; ROSA M. MEDA LARA <sup>2</sup>;

BERNARDO MORENO JIMÉNEZ <sup>3</sup>; PEDRO JUÁREZ RODRÍGUEZ <sup>4</sup>

(CENTRO UNIVERSITARIO DE LA CIÉNEGA. UNIVERSIDAD DE GUADALAJARA, Ocotlán, Jalisco, México)

### RESUMEN

**OBJETIVO:** conocer la estructura factorial del GHQ-12 en México y explorar su utilidad para identificar a personas que perciben tener problemas de salud. **METODO:** participaron 1093 adultos sin diagnóstico de enfermedades crónicas; 462 fueron varones, con una edad media de 32.2 años (DT=13.2) y 631 mujeres, con una edad media de 29.1 años (DT=12.0). Se realizaron comparaciones de correlaciones, medias y varianzas para explorar los efectos de la redacción de ítems en forma negativa versus positiva. Posterior al AFE, se compararon los índices de ajuste de los modelos multidimensionales, del unidimensional original y del unidimensional con 'efectos del método'. Se comparó la percepción general de salud y los puntajes del GHQ-12 para mujeres y hombres. **RESULTADOS:** el modelo unidimensional con 'efectos de método' (errores correlacionados) mostró el mejor ajuste a los datos. Los puntajes de ítems positivos, negativos y el total fueron más altos para participantes con problemas de salud, pero solo las varianzas e intervalos de confianza de los ítems positivos fueron consistentes y homogéneos. **CONCLUSION:** se apoya que el GHQ-12 es una medida unidimensional. Aplicado en población mexicana muestra una estructura factorial semejante a la de países orientales y occidentales. Se sugiere utilizar el puntaje de los ítems positivos del instrumento como un método general de tamizaje del bienestar/malestar psicológico.

**PALABRAS CLAVE:** Análisis factorial confirmatorio, efectos de redacción, estructura factorial, GHQ-12, México.

### ABSTRACT

**OBJECTIVES:** 1. To know the factor structure of the GHQ-12 in Mexican adult population. 2. To explore its usefulness to identify individuals who report health problems. **METHOD:** The study was carried out with a sample composed of 1093 adults without diagnosis of chronic diseases, divided into 462 males with an average age of 32.2 (DT = 13.2) and 631 females with an average age of 29.1 (DT = 12.0). In order to explore worded items effects positively and negatively, correlations, means, and variances were compared. Following the EFA, the adjustment of index of multidimensional models (the 'original' unidimensional model and the unidimensional with method effects) was compared. Also comparisons between self-reported health and GHQ-12 scores for both sexes were performed. **RESULTS:** The unidimensional model with 'method effects' obtained the best adjustment. Scores from positive, negative, and total items were higher for participants with self-reported health problems; however, only variances and confidence intervals with positive items were found homogeneous and consistent. **CONCLUSION:** GHQ-12 is confirmed as a unidimensional measure. Structural factor is similar to that of western and eastern countries when applied to Mexican population. Using the scores of GHQ-12 positive items as a screening method of psychological welfare/psychological is advised.

**KEY WORDS:** Confirmatory analysis, factor structure, GHQ-12, Mexico, wording effects.

1. Afiliado al Centro Universitario de la Ciénega. Universidad de Guadalajara, Ocotlán, Jalisco, México E-mail: psolis@mexis.com

2. Afiliado al Centro Universitario de Ciencias de la Salud, Guadalajara, Jalisco, México. E-mail: rosammeda2013@gmail.com

3. Afiliado a la Universidad Autónoma de Madrid, Madrid, España. E-mail: bernardo.moreno@gmail.com

4. Afiliado al Centro Universitario de Ciencias de la Salud, Guadalajara, Jalisco, México. E-mail: psic.pjr@gmail.com

## INTRODUCCIÓN

La salud es considerada uno de los bienes más importantes de los seres humanos, por lo que los países destinan cada vez una mayor cantidad de recursos a la mejora de los servicios sanitarios. Con ese objetivo se han realizado numerosos estudios dirigidos a analizar el estado de salud de las personas desde diversas perspectivas, entre los que destacan los que estudian la salud percibida como variable clave para analizar otros fenómenos y procesos, o bien para compararlo con otros indicadores de salud (Castro-Vázquez, Espinoza-Gutiérrez, Rodríguez-Contreras & Santos-Iglesias, 2007). La salud percibida se refiere a la valoración que tiene el individuo acerca de su salud, la cual es considerada un buen predictor de la esperanza de vida, la morbilidad y la utilización de servicios sanitarios (Idler & Benyamini, 1997). Existe evidencia que indica que los individuos con una mala salud percibida pueden utilizar con más frecuencia los servicios médicos y presentan un mayor número de ausencias en el trabajo, en comparación con aquellos con una buena percepción de salud (Miilunpalo, Vuori, Oja, Pasanen & Urponen, 1997).

La autopercepción del estado de salud aunque es un indicador subjetivo, representa una aproximación confiable del estado de salud general, así como un parámetro válido para la medición del estado de salud y un importante componente de la calidad de vida. La autopercepción buena o excelente acerca de la salud se relaciona con una mayor supervivencia, mientras que en su forma negativa se vincula con un riesgo mayor de mortalidad (Vladislavovna-Douvoba, Pérez-Cuevas, Reyes-Morales, 2008). De hecho, a través de estudios longitudinales se ha encontrado que la supervivencia se encuentra más relacionada con la salud subjetiva que con la objetiva (Tirado-Cosío & Moreno-Castillo, 2004).

Como parte de los intentos para medir el grado de salud percibida, diversos instrumentos han sido desarrollados, pero pocos con tan amplia utilización como el General Health Questionnaire (GHQ en adelante); este es un instrumento que fue diseñado en 1979 por David Goldberg en el Hospital Universitario de Manchester, Inglaterra, para el auto reporte del grado de bienestar general. Originalmente consta de 60 preguntas, que recorren 4 áreas psiquiátricas básicas: depresión, disfunción social, ansiedad, e hipocondría (Goldberg & Hillier, 1979). Si bien el nombre de este instrumento alude a la salud general, se reconoce que Goldberg lo elaboró para la detección de la severidad de disturbios psiquiátricos menores, es decir, la salud mental, así como para la valoración de algunas funciones intelectuales y fisiológicas (García-Viniegras, 1999).

Su vasta utilización a nivel internacional ha producido el surgimiento de adaptaciones en versiones reducidas de 30, 28 y 12 reactivos principalmente, todos de gran utilidad en distintos países. Una importante cantidad de las propuestas de adaptación de este instrumento provienen de poblaciones europeas y asiáticas (p. ej., Johnson, 2015; Li, Chung, Chui & Chan, 2009; Nagyova, et al., 2000; Riberio & Antunes, 2003). No obstante, países de habla hispana también han realizado adaptaciones con una gran diversidad de poblaciones (García-Viniegras & Rodríguez, 2007; Campo-Arias, 2007; Sánchez-López & Dresch, 2008).

Las adaptaciones del GHQ más ambiciosas son las versiones compuestas por 12 ítems; sin embargo, su estructura factorial no está todavía clara del todo pues los resultados de los análisis sobre la dimensionalidad de esta versión son contradictorios. Diversos estudios abogan por una solución monofactorial (p. ej., Villa, Zuluaga & Restrepo, 2013) en la que los 12 ítems describen una única dimensión de salud general o bienestar/malestar

psicológico, otros encuentran mayor coherencia en un modelo bifactorial (Cifre & Salanova, 2000), un factor de ansiedad y otro de depresión, mientras que otros tantos apuestan por una estructura de tres factores (Graetz, 1991; Rocha, Pérez, Rodríguez-Sanz, Borell & Obiols, 2011; Sánchez-López & Dresch, 2008).

Los estudios que pretenden aclarar la dimensionalidad de esta versión (GHQ-12) quizá se puedan comprender mejor al agruparlos en tres grupos. Aquellos que responsabilizan al método de análisis factorial elegido como responsable de los hallazgos. Por ejemplo, González e Ibáñez (2001) argumentaron que las divergencias entre los modelos de 1 y 2 factores podrían deberse a la utilización de métodos de Análisis Factorial Exploratorio (AFE) que tenderían a defender una solución monofactorial y los de Análisis Factorial Confirmatorio (AFC) que apoyarían una solución bifactorial o de tres factores.

Otro grupo es afín a las suposiciones de Rey, Abad, Barrada, Garrido y Ponsoda (2014) quienes exponen que los distintos esquemas de puntuación podrían explicar los resultados mixtos sobre la dimensionalidad. A este tenor, Urzúa, Caqueo-Urizar, Bargsted e Irrázaval (2015) refieren que el procedimiento de puntuación 0-1-1-1 con punto de corte 7 u 8, para los modelos de 2 y 3 factores presenta mejor ajuste, lo que sugeriría una relación entre el esquema de respuesta y el número de factores identificados en la estructura del instrumento. Urzúa et al. (2015) refieren que el modelo que muestra mejor ajuste entre todos es el de 3 factores con formato 0-1-1-1; sin embargo, el formato 0-1-2-3 con punto de corte 11 o 12, tiene indicadores aceptables y más estables, siendo un mejor formato para modelos bi y tridimensionales en población Iberoamericana. Estos resultados son coincidentes con el trabajo de Campo-Arias (2007) respecto a la puntuación 0-1-2-3 y 0-1-1-1 en modelos bifactoriales (ansiedad-depresión).

Un tercer grupo refiere que la redacción de los ítems que componen el GHQ-12, recordando que 6 ítems están escritos en forma positiva y 6 en forma negativa, es responsable de las contradicciones estructurales reportadas por las ambiguas categorías de respuesta a los ítems escritos de forma negativa, los cuales afectarían las respuestas de los individuos y por consecuencia la estructura factorial del cuestionario (Molina, Rodrigo, Losilla & Vives, 2014; Rey, et al., 2014; Smith, Fallowfield, Stark, Velikova & Jenkins, 2010). Al parecer, el estudio de Hankins (2008) en Inglaterra fue el pionero en señalar los efectos psicométricos de la redacción de ítems del GHQ-12 en forma negativa. En su estudio, demuestra que las varianzas de los ítems negativos son más grandes que las de los positivos; así como la frecuencia de respuestas a los ítems positivos se concentra en el punto 2 (en una escala Likert de 1 a 4 puntos), mientras que para los negativos se concentra en el 1 y el 2. Hankins replica los modelos más reconocidos previos a su estudio (unidimensional, bifactorial y tridimensional), encontrando los resultados usuales. Y presenta un cuarto modelo incorporando el efecto del método (unidimensional con errores correlacionados); este modelo presentó el mejor ajuste de todos los modelos, por lo que confirma que el GHQ-12 es unidimensional con sesgo en las respuestas (ítems negativos). En cuanto a la explicación conceptual, Hankins (2008) argumenta que lo ambiguo del sistema de calificación favorece las contradicciones estructurales y no la falta de atención, educación o la aversión al texto negativo.

Ye (2009) presenta una traducción del inglés al chino del GHQ-12 a una muestra de 348 universitarios para estudiar el problema de redacción de ítems y examinar la estructura factorial de dos modelos tradicionales (2 y 3 factores), y un modelo con un factor general y un factor de método (control de ítems negativos); todos los modelos mostraron un buen ajuste a los

datos, con la ventaja de que el modelo propuesto no presentaba los factores correlacionados. Ye (2009) encontró que las medias aritméticas de los ítems positivos era más altas que las de los negativos, y los ítems positivos presentaban asimetría y curtosis negativas y los ítems positivos lo contrario. Ye (2009) concluye que no se justifican factores correlacionados que, además, no proveen de una mejoría en el poder predictivo del GHQ-12 por lo que apoya la naturaleza unidimensional del cuestionario y sugiere aplicarlo tal cual es, pero utilizar solo los ítems positivos al calificarlo.

En otro estudio, Abubakar y Fischer (2012), en Kenia, replican el modelo original del GHQ-12 (unidimensional), el de Hankins (2008; unidimensional con errores correlacionados), la propuesta de Ye (un factor general y uno controlando la redacción negativa), el de dos factores (uno de ítems positivos y otro de negativos), y el tridimensional (propuesta de Graetz, 1991). En los índices de ajuste a los datos, el mejor ajuste lo presentó el modelo tridimensional; sin embargo, Abubakar y Fischer (2012) comentan que ese modelo separa a los ítems positivos y negativos y además los factores están correlacionados, afectando el poder predictivo del instrumento. Sugieren la utilidad del puntaje total como el más adecuado en términos de parsimonia, plausibilidad y utilidad. Sugieren también la opción de no considerar los ítems negativos al calificar el instrumento.

Aguado et al. (2012) exploran la estructura factorial y discriminante del GHQ-12 en mujeres españolas en fase postparto. En su estudio comparan el mejor ajuste de los datos de múltiples modelos, encontrando, nuevamente, que los modelos bi y tridimensionales no solo no sobresalen por su ajuste, si no que presentan correlaciones entre factores muy altas. Además prueban diferentes sistemas de calificación y no encuentran ventajas del sistema binario sobre el Likert. Confirman

que la estructura multidimensional es un artificio de estilos de respuesta (debida a los ítems negativos) y se suman a la evidencia a favor del GHQ-12 como medida unidimensional. Debido a que Aguado et al. (2012) solo estudiaron a mujeres y encontraron resultados semejantes a otros incluyendo a mujeres y hombres, la estructura factorial del GHQ-12, con efectos de método, parece adecuada para ambos sexos. En cuanto al poder discriminativo encontraron que solo el primer factor del modelo tridimensional discrimina y solo lo hace entre personas con diagnóstico o sin él; por lo que no es más eficiente que la puntuación total del GHQ-12.

Finalmente, otro estudio (Smith, Oluboyede, West, Hewison & House, 2013) compara los modelos propuestos y dos métodos de calificación, resultando el método Likert el más adecuado. Smith et al. (2013) utilizan datos de un estudio de Inglaterra sobre envejecimiento, por lo que toda la muestra era de 50+ años de edad. El modelo con mejor ajuste fue el unidimensional con errores correlacionados (Hankins, 2008). Utilizan también el método de agrupamiento para identificar si había dos tipos de respondientes en el GHQ-12, lo que confirmaron. Como sería de esperar el grupo de bajos puntajes calificó más bajo en todos los ítems comparado con el grupo con alto puntaje; pero en particular en los ítems negativos; y, además, la varianza de los ítems fue mayor para el grupo de alto puntaje, confirmando que el grupo de bajos puntajes mostraba poca varianza en sus ítems, particularmente en los negativos. Conceptualmente y a diferencia de los argumentos de Hankins (2008), Smith et al. (2013) argumentan que sus resultados con el grupo de altos puntajes, es decir con altos niveles de estrés, se debe a que ese grupo puede verse afectado en forma desproporcionada por los ítems negativos, explicando, quizás, las diferencias encontradas en distintas poblaciones.

Este estudio fue realizado con un doble propósito. El primero para conocer la

estructura factorial del GHQ-12 en población general de México. Debido al debate actual sobre la estructura factorial más adecuada para el GHQ-12, en el que se advierte que no tomar en cuenta el sesgo de los ítems negativos afecta los análisis estructurales y favorece la postulación de varios factores como auténticos constructos, siendo artificios del método de medida (Hankins, 2008), obliga necesariamente a examinar y comparar los modelos más populares. En el estudio se examinaron y compararon varios modelos: El modelo unidimensional original propuesto como medida de salud general o bienestar/malestar psicológico; el modelo bidimensional de factores correlacionados (p. ej., Cifre & Salanova, 2000); el modelo tridimensional (Graetz, 1991) y el modelo unidimensional con efectos del método (un factor con errores correlacionados de ítems negativos; Hankins, 2008). El segundo propósito fue explorar la utilidad del GHQ-12 para identificar a personas que perciben tener problemas de salud de las que no lo hacen. Un método para identificar la salud subjetiva es el de realizar una sencilla pregunta a los participantes sobre su estado de salud. Por ejemplo, Kinnunen et al. (2012), valoraron la salud subjetiva con una pregunta referente a cómo describirían los individuos su salud en el último año, y con el GHQ-12 como medida de distrés psicológico, encontrando relación entre ambos métodos. De igual manera, en este estudio se esperaría encontrar relación entre el puntaje del GHQ-12 y la pregunta sobre la salud de los participantes. Además, los análisis al respecto se planearon realizar por sexo. Esto porque datos epidemiológicos muestran que los hombres suelen presentar una mejor salud percibida con respecto a las mujeres pues ellas reportan con mayor frecuencia síntomas psíquicos y físicos (Castro-Vázquez, et al., 2007; Piko, 2000; Kaleta, Polańska, Dzionkowska-Zaborszcsyk, Hanke & Drygas, 2009).

Hasta donde sabemos este es el primer estudio que examina los modelos de

bondad de ajuste más populares y el modelo de efectos del método para conocer la estructura factorial de GHQ-12 en población mexicana y esperamos que los resultados contribuyan con datos de la estructura del GHQ-12 de importancia para el contexto latinoamericano. En este trabajo se espera contestar a las siguientes preguntas: 1.- ¿Muestra el GHQ-12 la estructura factorial encontrada en otras regiones del mundo (p. ej., Inglaterra, Kenia, China). 2.- ¿Es la estructura factorial del GHQ-12 afectada por los efectos de la redacción de ítems en forma negativa (efecto del método) en población mexicana? 3.- ¿Existe relación entre el puntaje del GHQ-12 elegido a partir del modelo con mejor bondad de ajuste y la pregunta sobre el estado de salud de los participantes?

## MÉTODO

### Participantes

La muestra es incidental y el total de participantes fue de 1093 adultos "relativamente sanos" sin diagnóstico de enfermedades crónicas, (estudiantes universitarios, profesores, empleados, obreros y amas de casa, residentes de los Estados de Jalisco y Veracruz en México). De entre ellos, 462 fueron varones, con una edad media de 32.2 años ( $DT=13.2$ ) y 631 mujeres, con una edad media de 29.1 años ( $DT=12.0$ ). Los porcentajes más altos indican que las características predominantes de la muestra eran solteros (63%), estudiantes de nivel superior (45%, Universidad), y sin problemas de salud (70%); concentrada la edad en grupos, el mayor porcentaje fue de menores de 40 años (73%).

### Instrumentos

*Cuestionario sociodemográfico.* Requiriendo edad, estado civil, escolaridad, y ocupación de los participantes; además, se les pidió que contestaran a la pregunta ¿padece algún problema de salud?

*Cuestionario General de Salud (GHQ-12)*. Diseñado por Goldberg y Williams (1988), es un cuestionario que consta de 12 reactivos que evalúan la salud auto percibida, el estado de bienestar general, especialmente en lo que se refiere a la presencia de ciertos estados emocionales. También pretende la evaluación de algunas de sus funciones intelectuales y fisiológicas, y la autovaloración que hace el individuo en el establecimiento y alcance de sus metas y propósitos en la vida y del enfrentamiento a las dificultades. Cada ítem presenta cuatro opciones de respuesta: mucho más de lo habitual, bastante más de lo habitual, no más de lo habitual, y no en lo absoluto. A cada ítem se dio una puntuación de cero a tres (0-1-2-3), en consecuencia, se obtienen puntuaciones totales que pueden oscilar entre 0 y 36. Procedimiento

Los instrumentos fueron convertidos a formato electrónico a través del software Survey Monkey, que incluía además una ficha socio demográfica y una carta de consentimiento informado. Posteriormente, el test fue enviado a la dirección de correo electrónico de potenciales participantes, adultos de los 18 a 60 años de edad que residían en los Estados de Jalisco y Veracruz (México). Los datos fueron obtenidos en el período comprendido de octubre 2012 a marzo del 2013.

El proyecto fue aprobado por el Consejo de Ética de la Universidad de Guadalajara. Además fueron tomados en consideración la Declaración de Helsinki de 2004 y el Código Ético del Psicólogo (Sociedad Mexicana de Psicología, 2007), cuidando la integridad de los participantes y haciendo un uso confidencial de los datos.

#### Análisis de datos

Todos los análisis estadísticos se realizaron con el software StatSoft (2011). Se eligió el método de calificación Likert (escala de 0 a 3) porque las evidencias en los estudios estructurales apoyan este método (p. ej.,

Hankins, 2008); para facilitar la comprensión a la numeración original de los ítems se les agregó una letra para indicar su redacción en forma positiva (P) o negativa (N). Los análisis de datos consistieron en analizar las varianzas de los ítems con intervalos de confianza (IC) de 95% con la intención de corroborar si los ítems negativos mostraban mayor varianza que los positivos; además, la obtención de la matriz de correlaciones Pearson de los ítems, incluyendo medias, desviaciones, asimetría y curtosis fueron incluidas para explorar la naturaleza de los tipos de respuesta ante ítems positivos *versus* negativos (Ye, 2009) dados los argumentos a favor de la unidimensionalidad del GHQ tanto como a favor de la multidimensionalidad. Los modelos evaluados se pueden dividir en tres tipos de acuerdo a varios autores (Aguado, et al., 2012): 1) el modelo unidimensional original y el modelo unidimensional con efectos del método (errores correlacionados para los ítems parafraseados en forma negativa; Hankins, 2008); 2) modelo bifactorial con factores correlacionados (p. ej., Cifre & Salanova, 2000); y 3) modelo trifactorial con factores correlacionados (p. ej., Graetz, 1991). Puesto que una de las críticas que ha dado lugar a la ejecución de análisis estructurales, y que ha justificado el análisis comparativo de ítems parafraseados en forma negativa *versus* positiva, ha sido que el análisis factorial exploratorio (AFE) genera dos factores (o tres) que separan los ítems positivos de los negativos, se realizó un AFE con rotación oblicua (Oblimin) para corroborar este caso. Se realizaron varios análisis factoriales confirmatorios (AFC), posteriores al AFE. Siguiendo las sugerencias de la literatura, la bondad de ajuste para cada modelo incluyó varios parámetros: Sattora-Bentler Chi cuadrada; el CFI, donde valores adecuados se consideran en .96 o más, y el ajuste perfecto en 1; el NNFI y el NFI, donde los mínimos valores adecuado son de .95; y el RMSEA, donde valores menores de 0.08

indican adecuado ajuste. Se obtuvieron los alfa de Cronbach de todos los ítems para conocer si la fiabilidad de los ítems parafraseados en negativo era diferente, como se ha sugerido (Hatkins, 2008). En relación a la comparación de los puntajes del GHQ-12, según la percepción general de salud, se utilizó la prueba *t* de Student para muestras independientes asumiendo homogeneidad de varianzas. Se compararon los ítems positivos, los negativos y el total, en mujeres, en hombres y en la muestra total; se incluyeron las varianzas y los IC (95%) para corroborar el comportamiento estadístico diferencial por tipo de ítems.

## RESULTADOS

### Análisis Exploratorios

En la Tabla 1 se presentan las correlaciones entre ítems, las medias y desviaciones típicas, la asimetría y la curtosis de cada ítem. Las correlaciones fueron resaltadas para facilidad en su lectura y se puede observar que, entre ítems negativos (N2,

N5, N6, N9, N10, N11), éstas iban de .23 a .61 ( $M = .43$ ), y entre los positivos (P1, P3, P4, P7, P8, P12), de .40 a .58 ( $M = .34$ ). Las medias de los ítems positivos iban de 1.10 a 1.43 ( $M = 1.20$ ), mientras que para los ítems negativos de 0.38 a 1.41 ( $M = 0.92$ ); Las curtosis fueron en su mayoría negativas y las asimetrías positivas, sin distinción entre ítems positivos y negativos; estos resultados indican que la muestra estaba conformada por personas que se consideraban psicológicamente sanas. Sin embargo, en la Tabla 1 se puede observar que la asimetría y la curtosis de los ítems negativos se desvió de la normalidad, en particular los ítems N10 y N11, se desviaron mucho de la normalidad.

Se obtuvieron las varianzas y el intervalo de confianza de 95% para los dos tipos de ítems. En la Tabla 2 claramente se observa que las varianzas de los ítems negativos fueron superiores a las de los positivos, pero se observa también que los rangos de las varianzas y de los intervalos de confianza se traslapan entre tipos de ítems.

TABLA 1.  
Correlaciones, medias, asimetrías y curtosis de los ítems del GHQ-12.

	P1	N2	P3	P4	N5	N6	P7	P8	N9	N10	N11	P12
P1. Concentración	-											
N2. Pérdida de sueño	.04	-										
P3. Ser útil	<b>.43</b>	.01	-									
P4. Tomar decisiones	<b>.42</b>	.03	<b>.56</b>	-								
N5. Tensión	.02	<b>.43</b>	.03	.09	-							
N6. Vencer dificultades	.10	<b>.44</b>	.09	.10	<b>.43</b>	-						
P7. Disfrutar la vida	<b>.44</b>	.07	<b>.48</b>	<b>.49</b>	.09	.14	-					
P8. Afrontar problemas	<b>.40</b>	.01	<b>.50</b>	<b>.53</b>	.07	.08	<b>.58</b>	-				
N9. Sentirse triste	.10	<b>.45</b>	.09	.07	<b>.39</b>	<b>.51</b>	.19	.04	-			
N10. Pérdida de confianza	.13	<b>.35</b>	.19	.18	<b>.31</b>	<b>.52</b>	.23	.16	<b>.61</b>	-		
N11. No valer nada	.12	<b>.29</b>	.23	.21	<b>.23</b>	<b>.42</b>	.24	.19	<b>.49</b>	<b>.68</b>	-	
P12. Ser feliz	<b>.41</b>	.04	<b>.45</b>	<b>.46</b>	.01	.15	<b>.53</b>	<b>.47</b>	.18	.21	.25	-
Media	1.43	1.17	1.22	1.12	1.41	0.97	1.17	1.18	1.02	0.61	0.38	1.10
DT	0.78	0.97	0.84	0.84	0.94	0.93	0.87	0.84	0.94	0.88	0.77	0.86
Asimetría	-0.43	0.44	0.00	0.11	0.23	0.66	0.05	-0.00	0.67	1.34	2.11	0.21
Curtosis	-0.59	-0.77	-0.86	-0.92	-0.84	-0.48	-1.00	-0.95	-0.40	0.82	3.69	-0.85

TABLA 2.  
 Varianzas (95% Intervalo de confianza) de los ítems positivos (P) y negativos (N).

Ítems P	Varianza	95% IC para la varianza
1P	0.61	1.38-1.47
3P	0.70	1.17-1.27
4P	0.71	1.07-1.17
8P	0.71	1.13-1.23
12P	0.74	1.05-1.16
7P	0.76	1.12-1.22
Ítems N		
11N	0.59	0.33-0.42
10N	0.78	0.56-0.66
6N	0.88	0.91-1.03
5N	0.89	1.35-1.46
9N	0.89	0.96-1.07
2N	0.94	1.16-1.23

Previo al análisis factorial exploratorio (AFE), se realizaron la prueba de esfericidad de la muestra de Bartlett y la prueba de adecuación de la muestra de Kayser-Meier-Olkin (KMO). Los resultados de la prueba de Bartlett ( $X^2(66) = 4825.130$ ,  $p < .001$ ) y de KMO (.86) indican que la muestra es adecuada para los análisis propuestos. En el análisis, la gráfica (no presentada aquí) de los valores propios sugirió que dos, o tal vez tres, eran los

factores necesarios para obtener una solución factorial simple del GHQ, con normalización Kaiser ambas, y aceptando factores con valores propios mayores que 1; la solución factorial se presenta en la Tabla 3. Como se esperaba, los factores corresponden a los más reportados tradicionalmente; el primero agrupa a los ítems positivos y, el segundo, a los ítems negativos. El total de varianza explicada por estos factores fue de 55.8%.

TABLA 3.  
 Solución del AFE con rotación oblicua.

GHQ-12 Ítems	Factores	
	1	2
P1. Concentración	<b>.581</b>	.107
N2. Pérdida de sueño	.016	<b>.581</b>
P3. Ser útil	<b>.708</b>	.116
P4. Tomar decisiones	<b>.730</b>	.077
N5. Tensión	.033	<b>.534</b>
N6. Vencer dificultades	.148	<b>.696</b>
P7. Disfrutar la vida	<b>.725</b>	.210
P8. Afrontar problemas	<b>.735</b>	.069
N9. Sentirse triste	.160	<b>.765</b>
N10. Pérdida de confianza	.282	<b>.762</b>
N11. No valer nada	.313	<b>.647</b>
P12. Ser feliz	<b>.666</b>	.187

Nota: se presentan valores  $> .50$  resaltados.

#### Análisis Estructurales

En la Tabla 4 se presentan los índices de bondad de ajuste de cada uno de los modelos y se observa que de todos los modelos tradicionalmente reportados, destaca con un ajuste relativamente aceptable el modelo tridimensional. Es decir, el valor de RMSEA ( $< .08$ ) y el de CFI

(.95), son adecuados, y los valores de NFI (.94) y NNFI (.94) son muy cercanos a lo esperado. Sin embargo, el modelo unidimensional con efectos del método (errores correlacionados en los ítems negativos) muestra índices de bondad de ajuste claramente mejores.



TABLA 4.  
Índices de ajuste de los modelos factoriales del GHQ-12.

Modelo	X <sup>2</sup>	df	RMSEA (90% IC)	CFI	NFI	NNFI
Unidimensional	2272.788	54	0.24	0.54	0.53	0.43
Bidimensional	542.874	54	0.098	0.90	0.89	0.88
Tridimensional	275.557	51	0.065	0.95	0.94	0.94
Unidimensional con errores correlacionados	147.968	39	0.051	0.98	0.97	0.96

Consistencia interna

El coeficiente alfa de Cronbach para los ítems positivos fue de .85, no mejoraba al retirar alguno de los ítems, y su media aritmética fue de 7.24 (DT = 3.80; varianza = 14.47). Y, para los ítems negativos, el coeficiente fue de .82, no mejoraba al retirar alguno de los ítems (aunque retirar el ítem N5 no afectaba la fiabilidad), y su media aritmética fue de 5.57 (DT = 3.97; varianza = 15.81). Para la suma de todos los ítems el coeficiente fue de .81; la media aritmética fue de 12.8 (DT = 5.99; varianza = 35.88).

Percepción de salud

Del total de la muestra, el 30.3% indicó presentar problemas de salud; en las mujeres el 46% y en los hombres el 40%. En la Tabla 5 se presentan los porcentajes de cada tipo de problema de salud en función del número de personas que reportaron tener ese problema.

TABLA 5.  
Frecuencia de problemas de salud reportados por sexo.

Problema de salud	Mujeres		Hombres	
	n	%	n	%
Cardiovascular	12	6.0	11	8.3
Cáncer	2	1.0	3	2.3
Obesidad	55	27.6	20	15.2
Insomnio	40	20.1	31	23.5
Diabetes	16	8.0	26	19.7
Sistema circulatorio	7	3.5	1	0.8
Gastritis	64	32.2	33	25.0
Otras	45	22.6	22	16.7

Nota: mujeres (n = 199); hombres (n = 132).

En la Tabla 6 se presentan las diferencias entre los puntajes de los ítems positivos, los negativos y la sumatoria de ambos, según la percepción de salud de las mujeres, los hombres, y el total de ambos. Las pruebas *t* con estimados de varianza homogéneas, indicaron diferencias significativas en todas las comparaciones y con mayor probabilidad al juntar las muestras de hombres y mujeres. Las muestras que reportaron tener problemas de salud tuvieron puntajes más elevados en los tres tipos de puntajes, indicativos de mayor malestar psicológico. En la Tabla 6 se presentan también las varianzas y los intervalos de confianza (IC 95%) por tipo de ítems y para el total; se puede observar que las varianzas y los IC son consistentes para los ítems positivos en mujeres, hombres y en el total de la muestra, pero no para los ítems negativos; en estos últimos se observa que la varianza se modifica según se trate de mujeres, hombres, o el total. Al sumar ambos tipos de ítems, se observa para los hombres y en particular para el total de la muestra que los IC no incluyen el valor de la varianza respectiva.

TABLA 6.

Comparación de puntajes de ítems positivos (IP), negativos (IN) y del total (GHQ), según la percepción de salud de mujeres, hombres y ambos.

Problema de salud	Mujeres				<i>t</i> (g.l. = 629)	<i>p</i> =	Varianza	IC 95%
	Sí ( <i>n</i> = 199)		No ( <i>n</i> = 432)					
	<i>M</i>	<i>DT</i>	<i>M</i>	<i>DT</i>				
IP	8.03	3.96	7.21	3.68	2.53	.05	1.15	0.18-1.45
IN	6.77	4.45	5.68	3.85	3.14	.01	1.33	0.40-1.76
GHQ	14.8	6.34	12.9	5.75	3.74	.000	1.21	0.90-1.90
Problema de salud	Hombres				<i>t</i> (g.l. = 460)	<i>p</i> =	Varianza	IC 95%
	Sí ( <i>n</i> = 132)		No ( <i>n</i> = 330)					
	<i>M</i>	<i>DT</i>	<i>M</i>	<i>DT</i>				
IP	7.82	3.97	6.59	3.69	3.16	.01	1.16	0.46-1.99
IN	5.73	3.75	4.66	3.72	2.79	.01	1.01	0.31-1.82
GHQ	13.5	6.01	11.2	5.67	3.87	.000	1.12	1.13-3.46
Problema de salud	Ambos				<i>t</i> (g.l. = 1091)	<i>p</i> =	Varianza	IC 95%
	Sí ( <i>n</i> = 331)		No ( <i>n</i> = 762)					
	<i>M</i>	<i>DT</i>	<i>M</i>	<i>DT</i>				
IP	7.94	3.96	6.94	3.69	3.91	.000	1.14	0.51-1.49
IN	6.36	4.21	5.24	3.82	4.30	.000	1.21	0.60-1.62
GHQ	14.3	6.23	12.2	5.77	5.44	.000	1.16	1.35-2.88

## DISCUSIÓN

Hasta donde los autores sabemos, este es el primer estudio en población mexicana que compara los modelos factoriales del GHQ-12 más reconocidos en la literatura especializada. Se compararon los índices de ajuste del modelo original unidimensional, del bidimensional (p.ej., Campo-Arias, 2007; Rocha, et al., 2011; Villa, et al., 2013), del tridimensional (Graetz, 1991), y el unidimensional con efectos del método (Hankins, 2008), con el propósito de corroborar hallazgos recientes que cuestionan la multidimensionalidad del GHQ-12 (Abubakar & Fisher, 2011; Aguado, et al., 2012; Molina, et al., 2014; Smith, et al., 2013; Ye, 2009). La ausencia de una estructura factorial común a través de estudios dificulta la utilización del GHQ-12 como instrumento de tamizaje de la salud mental general y genera incertidumbre en la toma de decisiones en la práctica clínica.

Hankins (2008) argumentó y demostró que los análisis factoriales que apoyan la naturaleza multidimensional del GHQ-12 no toman en cuenta que los ítems parafraseados en formato negativo podían

generar divisiones espurias de factores (dos o tres), debidas a efectos de las respuestas al formato negativo. En varios estudios se ha demostrado un comportamiento estadístico diferencial entre tipos de ítems del GHQ-12. En este trabajo encontramos que, para los ítems en formato negativo (IN), los rangos de las correlaciones y de las medias fueron más amplios, y las correlaciones fueron más fuertes que para los ítems en formato positivo (IP). También encontramos que, para ambos tipos de ítems, la mayoría de las curtosis eran negativas y las asimetrías positivas. Esto indica que las respuestas se acumularon en los puntos 0 y 1 para los IN y en el 1 para los IP, como se esperaba. En cuanto a la comparación de varianzas entre IP e IN se demostró claramente que los IN muestran mayor varianza que los IP. La mayoría de estos resultados son semejantes a los de otros autores, confirmando un comportamiento diferencial por tipo de ítem, caracterizado por diferencias en las varianzas entre IP e IN debido al prejuicio en las respuestas a los IN (Hankins, 2008; Smith, et al., 2013; Ye, 2009).

De importancia particular parece el hallazgo referente a dos ítems negativos (N10 y N11), que comúnmente son los únicos que saturan en el tercer factor de la solución tridimensional (Graetz, 1991), que algunos autores han descartado de las soluciones bifactoriales (p. ej., N11; Villa, et al., 2013), o han encontrado baja saturación en las soluciones de tres factores (Sánchez-López & Dresch, 2008), y que mostraron una clara desviación de la normalidad por su asimetría y curtosis en este estudio; además, en la comparación de varianzas entre tipo de ítems se encontró traslape entre los intervalos de confianza debido al comportamiento de las respuestas a los ítems N10 y N11. Por ello, nos referiremos a estos ítems más adelante.

En este trabajo los análisis factoriales confirmatorios apoyaron que el modelo unidimensional con efectos del método presenta el mejor ajuste entre los modelos evaluados y el modelo unidimensional original el ajuste más inadecuado; hallazgo común en varios estudios (Hankins, 2008; Smith, et al., 2013; Ye, 2009). También se obtuvieron índices adecuados para el modelo bidimensional; ese análisis fue congruente con la solución más simple del análisis factorial exploratorio. Aunque los índices de ajuste del modelo bidimensional fueron claramente mejores que los del unidimensional original, el RMSEA fue inadecuado. Como hemos visto, la solución factorial bidimensional ha sido propuesta como una adecuada, pero, entre otros problemas, los factores encontrados separan los IP y los IN, indicando un artificio del método de medida. Los estudios de las propiedades de las pruebas psicológicas demuestran que incluir ítems en formato negativo, aunque común y aceptado como una medida de control de la aquiescencia ante los ítems positivos, favorece la división de factores espurios que no debiesen interpretarse como constructos verdaderos (p. ej., Horan, DiStefano & Motl, 2003). En cuanto al modelo tridimensional (Graetz, 1991), encontramos que los índices de

ajuste fueron los más adecuados entre los modelos tradicionales, resultado común en varios estudios (Hankins, 2008; Smith, et al., 2013). Pero algunos estudios han encontrado que ese modelo muestra mejor ajuste que todos los modelos analizados (Ye, 2009), incluyendo dos modelos con efectos del método (Abubakar & Fischer, 2012). Sin embargo, en los modelos de dos y tres factores, se ha hallado que los factores están fuertemente correlacionados afectando el poder predictivo y de discriminación (Abubakar & Fischer, 2012; Aguado et al., 2012).

En cualquier caso, la mayoría de los estudios que han comparado los índices de ajuste de los modelos analizados en este trabajo reconocen que la decisión sobre el modelo adecuado del GHQ-12 no solo depende de los mejores índices de ajuste sino de otras consideraciones. Por ejemplo, Abubakar y Fischer (2012) al igual que Ye (2009) argumentan que el modelo unidimensional propuesto por Hankins (2008) es el más parsimonioso con el constructo original y con el propósito del instrumento como herramienta de tamizaje de la salud mental. Otra consideración es la practicidad al utilizar un solo puntaje ya que un puntaje unitario es más parsimonioso y suficiente para tamizaje. Pero ante la problemática que representa la redacción de ítems en forma negativa, se ha sugerido, entre otras alternativas, aplicar todo el GHQ-12, pero calificar solo los ítems positivos. Estudios futuros podrían comparar cuidadosamente esta opción y la de redactar los ítems negativos en forma positiva.

Por otra parte, en este trabajo se comparó la percepción subjetiva de salud con dos parámetros, el GHQ-12 y una pregunta sobre el estado de salud, en mujeres, en hombres y en el total de la muestra. Kinnunen et al. (2012) estudiando salud y personalidad, utilizaron un procedimiento semejante pero los participantes describían su salud en el

último año, y se utilizó el puntaje total del GHQ-12 como medida de distrés psicológico; encontraron diferencias significativas por tipo de personalidad en estos indicadores subjetivos. Esto es relevante para el presente estudio porque aquí se exploró si los IP serían suficientes para identificar problemas de salud en general. Las comparaciones de puntajes en el instrumento *versus* la respuesta ante la pregunta del estado de salud, indicaron que bastaría utilizar los IP para hallar un resultado semejante al encontrado con los IN o con la suma total de puntajes como en otros estudios (Kinnunen, et al., 2012; Rocha, et al., 2011). La mayor significancia de las comparaciones con IN o al sumar los puntajes de IP e IN, como vimos, es un artificio de las varianzas y no un indicador de diferencias en constructos como los de ansiedad y depresión que, con base en los datos, parecen sin fundamento. Como apoyo para la interpretación de los resultados, en este estudio obtuvimos la fiabilidad por tipo de ítems y por su sumatoria porque se ha argumentado que la inclusión de los IN estaría subestimando o sobreestimando la fiabilidad del GHQ-12 (Hatkins, 2008). Congruente con los supuestos, todos los coeficientes alfa mostraron fiabilidad aceptable (Hernández, Fernández & Baptista, 2010), pero la fiabilidad más alta se encontró para los IP. Todos estos resultados apoyan la noción referente a que bastaría con solo seis ítems, los IP, para obtener resultados semejantes que con 12, y que los IP pueden ser suficientes para un instrumento de tamizaje de la salud mental en general.

La interpretación conceptual de los hallazgos sobre el modelo unidimensional con 'efectos del método' ha variado entre un par de autores que se han referido al tema. Hankins (2008) argumentó que el prejuicio en las respuestas ante los IN por parte de población general no se debe a falta de atención o educación de la población, o al rechazo a la redacción en forma negativa, sino más bien al formato ambiguo de

respuestas que se presenta para indicar la ausencia de un estado emocional negativo. Así para indicar que "no se tiene" ese estado emocional, las personas enfrentan dos opciones de respuesta: no en lo absoluto (*not at all*) y no más de lo habitual (*no more than usual*); por lo que para los individuos sin problemas psiquiátricos esto representa una ambigüedad que incrementa la varianza de los IN. En contraste, Smith et al. (2013), al dividir su muestra en dos grupos, uno con puntajes altos y otro con bajos en el GHQ-12, hallaron que las varianzas de los IN del primer grupo fueron mayores que las de los IP; y argumentaron que el grupo con altos puntajes, o con altos niveles de estrés psicológico, podía verse afectado en forma desproporcionada por los IN; y que esos niveles de estrés psicológico probablemente afectan la estructura del GHQ-12 y puedan explicar las diferencias encontradas en poblaciones diversas. Aunque los datos de este estudio solo permiten algunas suposiciones sobre el tema, creemos que su mención es necesaria. Estamos de acuerdo con Hankins (2008), y nuestros datos así lo muestran, en que la mayoría de la población general posiblemente se vea "obligada" a responder ante las dos opciones de respuesta que niegan el estado emocional, aunque no se perciban muy claras. Por ejemplo, en este trabajo las respuestas a dos ítems parafraseados en formato negativo, a los cuales nos referimos antes, son el N10 (¿Ha perdido confianza en sí mismo?) y el N11 (¿Ha sentido que usted no vale nada?). Parece poco razonable responder que uno ha perdido la confianza en sí mismo 'más de lo habitual'; o, peor aún, que ha sentido que uno no vale nada 'más de lo habitual', en este último ítem, la gran mayoría de la población contestó 'no en lo absoluto'. Por ello, coincidimos con Smith et al. (2013) en cuanto a que las respuestas pueden estar reflejando estrés psicológico ante la ambigüedad de las opciones disponibles. Además, ítems como los mencionados, referentes de la confianza y la estima en uno mismo, probablemente

reflejan actitudes o creencias culturales de la población estudiada. Por supuesto demostrarlo en población general y psiquiátrica puede tener importantes implicaciones para la investigación y la práctica.

La administración de instrumentos no adaptados culturalmente ni validados puede producir sesgos en la información recabada impactando en los diagnósticos, en las decisiones en la terapia individual, en los datos epidemiológicos e incluso en el diseño y puesta en marcha de políticas públicas (Ramada-Rodilla, Serra-Pujadas & Delclós-Clanchet, 2013). Siendo el GHQ-12 uno de los instrumentos preferidos para tamizaje de trastornos menores en población general y para toma de decisiones en múltiples escenarios clínicos o no, demostrar con continuos estudios su estructura factorial es fundamental. En este estudio se replicaron los resultados de estudios occidentales y orientales (p. ej., China, Inglaterra, Kenia, España), confirmando que los ítems negativos y la suma del total de ítems del GHQ-12 es innecesaria para considerar al instrumento una medida de tamizaje del bienestar/malestar psicológico (González e Ibañez, 2001) de la población general.

Finalmente, el estudio hace una contribución significativa al favorecer la comprensión de la estructura factorial del GHQ-12 en un nuevo escenario cultural, que puede ser de importancia para el contexto latinoamericano en particular. Sin embargo, una limitación es que no provee de datos para identificar punto de corte del GHQ. Estudios futuros podrían incluir muestras con y sin diagnósticos clínicos para satisfacer esta necesidad, pero considerando al GHQ como medida unidimensional, parafraseando los ítems negativos en formato positivo o descartándolos del puntaje total.

## REFERENCIAS

- Abubakar, A., & Fischer, R. (2012). The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress and Health*, 28, 248–254.
- Aguado, J., Campbell, A., Ascaso, C., Navarro, P., Garcia-Esteve, L., & Luciano, J. V. (2012). Examining the factor structure and discriminant validity of the 12-Item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment*, 19(4) 517–525.
- Campo-Arias, A. (2007). Cuestionario general de salud-12: análisis de factores en población general de Bucaramanga, Colombia. *IATREIA*, 20(1), 29-36.
- Castro-Vázquez, A., Espinosa-Gutiérrez, I., Rodríguez-Contreras, P., & Santos-Iglesias, P. (2007). Relación entre el estado de salud percibido e indicadores de salud en la población española. *International Journal of Clinical and Health Psychology*, 7(3), 883-898.
- Cifre, E., & Salanova, M. (2000). Validación factorial del "General Health Questionnaire" (GH1-12) mediante un análisis factorial confirmatorio. *Revista de Psicología de la Salud*, 12(2), 75-89.
- García-Viniegras, C. R. V., & Rodríguez, G. (2007). Calidad de vida en enfermos crónicos. *Revista Habanera de Ciencias Medicas*, 6(4), 1-9.
- García-Viniegras, C. R. V. (1999). Manual para la utilización del cuestionario de salud general de Goldberg. Adaptación cubana. *Revista Cubana de Medicina General Integral*, 15(1), 88-97.
- Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, 9, 139-145.
- Goldberg, D. P., & Williams, P. (1988). A user's guide to the General Health Questionnaire. Windsor, UK: NFER-Nelson.

- González, M., & Ibáñez, I. (2001). Cuestionario de Salud General (GHQ-12): comparación de dos modelos factoriales. *Psiquiatría.com*, 5(1). Recuperado de <http://www.psiquiatria.com/revistas/index.php/psiquiatriacom/article/viewFile/533/512>
- Graetz, B. (1991). Multidimensional properties of the General Health Questionnaire. *Social Psychiatry and Psychiatric Epidemiology*, 26(3), 132-138.
- Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, 4(10). Recuperado de <http://www.cpementalhealth.com/content/4/1/10>
- Hernández, R., Fernández, C., & Baptista, P. (2010). *Metodología de la investigación*. (5ta. ed.). Perú: McGraw-Hill.
- Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodologica I artifact or response style? *Structural Equation Modeling*, 10(3), 435-455.
- Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38, 21-37.
- Johnson, L. R. (2015). Searching for a screener: Examination of the factor structure of the General Health Questionnaire in Malaysia. *International Perspectives in Psychology: Research, Practice, Consultation*, 4(2), 111-127.
- Kaleta, D., Polańska, K., Dzionkowska-Zaborszcsyk, E., Hanke, W., & Drygas, W. (2009). Factors influencing self-perception of health status. *Central European Journal of Public Health*, 17(3), 122-127.
- Kinnunen, M. L., Metsäpelto, R. L., Feldt, T., Kokko, K., Tolvanen, A., Kinnunen, U., Leppänen, E., et al. (2012). Personality profiles and health: Longitudinal evidence among Finnish adults. *Scandinavian Journal of Psychology*, 53, 512-522. <http://dx.doi.org/10.1111/j.1467-9450.2012.00969.x>
- Li, H. C., Chung, O. K., Chui, M. L., & Chan, S. L. (2009). Factorial structure of the Chinese version of the 12-item General Health Questionnaire in adolescents. *Journal of Clinical Nursing*, 18(23), 3253-3261.
- Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., & Urponen, H. (1997). Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology*, 50, 517-528.
- Molina, J. G., Rodrigo, M. F., Losilla, J. M., & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological Assessment*, 26(3), 1031-1037.
- Nagyova, I., Krol, B., Szilasiova, A., Stewart, R. E., van Dijk, J. P., & van den Heuvel, W. (2000). General Health Questionnaire-28: psychometric evaluation of the Slovak version. *Studia Psychologica*, 42(4), 351-361.
- Penninkilampi-Kerola, V., Miettunen, J., & Ebeling, H. (2006). A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. *Scandinavian Journal of Psychology*, 47, 431-440.
- Piko, B. (2000). Health-related predictors of self-perceived health in a student population. The importance of physical activity. *Journal of Community Health*, 25(2), 125-137.
- Ramada-Rodilla, J. M., Serra-Pujadas, C., & Delclós-Clanchet, G. L. (2013). Adaptación cultural y validación de cuestionarios de salud: revisión y

- recomendaciones metodológicas. *Salud Pública de México*, 55(1), 55-66.
- Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., & Ponsoda, V. (2014). The impact of ambiguous response categories on the factor structure of the GHQ-12. *Psychological Assessment*, 26(3), 1021-1030.
- Riberio, J. L., & Antunes, S. (2003). Contribuição para o estudo de adaptação do Questionário De Saúde Geral de 28 itens (General Health Questionnaire GHQ28) *Revista Portuguesa de Psicossomática*, 5(1), 37-45.
- Rocha, K. B., Pérez, K., Rodríguez-Sanz, M., Borell, C., & Obiols, J. E. (2011). Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española. *International Journal of Clinical and Health Psychology*, 11(1), 125-139.
- Sánchez-López, M. P., & Dresch, V. (2008). The 12-item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema*, 20(4), 839-843.
- Smith, A. B., Fallowfield, L. J., Stark, D. P., Velikova, G., & Jenkins, V. (2010). A Rasch and confirmatory factor analysis of the General Health Questionnaire (GHQ-12). *Health and Quality of Life Outcomes*, 8(45). Recuperado de: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2873594/>
- Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress. *Quality of Life Research*, 22, 145-152.
- Sociedad Mexicana de Psicología. (2007). *Código Ético del Psicólogo*. México: Trillas.
- StatSoft (2011). *Statistica. Data analysis, predictive data mining, quality control, web based analytics*. Tulsa, OK: StatSoft.
- Tirado-Cosío, F., & Moreno-Castillo, Y. (2004). Autopercepción de salud en un grupo de ancianos mexicanos con el uso del perfil de salud de Nottingham. *Archivo Medicina Familiar*, 6(1), 20-22.
- Urzúa, A., Caqueo-Úrizar, A., Bargsted, M. & Irrázaval, M. (2015). ¿Afecta la forma de puntuación la estructura factorial de GHQ-12? *Cadernos de Saúde Pública*, 31(6), 1305-1312.
- Villa, I. C., Zuluaga, C., & Restrepo, L. F. (2013). Propiedades psicométricas del Cuestionario de Salud General de Goldberg GHQ-12 en una institución hospitalaria de la ciudad de Medellín. *Avances en Psicología Latinoamericana*, 31(3), 532-545.
- Vladislavovna-Douvoba, S., Pérez-Cuevas, R., Reyes-Morales, H. (2008). Autopercepción del estado de salud en climatéricas derechohabientes del Instituto Mexicano del Seguro Social. *Salud Pública de México*, 50(5), 390-396.
- Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of negative wording effects. *Personality and Individual Differences*, 46, 197-201.