

ARTÍCULO ORIGINAL

Validez de constructo y consistencia interna de tres estructuras factoriales y dos sistemas de puntuación del cuestionario de salud general de 12 ítems

Miguel Simancas-Pallares¹, Katherine Margarita Arrieta², Luisa Leonor Arévalo³

¹ Departamento de Investigación, Facultad de Odontología, Universidad de Cartagena, Cartagena de Indias, Colombia

² Departamento de Odontología Preventiva y Social, Facultad de Odontología, Universidad de Cartagena, Cartagena de Indias, Colombia

³ Departamento de Medicina Oral y Cirugía, Facultad de Odontología, Universidad de Cartagena, Cartagena de Indias, Colombia

Introducción. El análisis de las propiedades psicométricas de un instrumento reviste importancia para el control de sesgos en la medición.

Objetivo. Comparar la validez de constructo y la consistencia interna de tres estructuras factoriales del cuestionario de salud general de 12 ítems.

Materiales y métodos. Se hizo un estudio de validación en una muestra anidada de 483 estudiantes de odontología de Cartagena, Colombia. Además de las preguntas del instrumento, se hicieron otras sobre aspectos personales. Se evaluó la validez de constructo de la estructura unidimensional, de la unidimensional con corrección del sesgo de respuesta, y de la de dos y tres factores y sistemas de puntuación del tipo de la escala de Likert (0-1-2-3) y dicotómico (0-0-1-1). La validez se determinó con un análisis factorial confirmatorio y, la consistencia interna, con el coeficiente alfa de Cronbach para la escala de Likert y la fórmula 20 de Kuder-Richardson para el método dicotómico empleando los programas Stata™, versión 13.2 (Stata Corp, USA) y Mplus™, versión 7.11 (Muthen and Muthen, USA).

Resultados. Los índices de ajuste revelaron que el mejor modelo era el unidimensional con corrección del sesgo de respuesta y puntuación dicotómica (grados de libertad=36; ji al cuadrado=52,432; raíz cuadrada del error medio de aproximación=0,03; IC_{90%} 0,008-0,048; índice comparativo de ajuste=0,982; índice de Tucker-Lewis=0,966). La consistencia interna fue de 0,70.

Conclusiones. El ajuste del modelo obtenido en este estudio permitió controlar el sesgo de medición y, así, garantizar la validez externa de los resultados del cuestionario de salud general de 12 ítems para la evaluación de las condiciones mentales en esta población. Las propiedades psicométricas de las escalas deben evaluarse críticamente antes del análisis de los resultados.

Palabras clave: estudios de validación; análisis factorial; sesgo (epidemiología); psicometría; salud mental; estudiantes de odontología.

doi: <https://doi.org/10.7705/biomedica.v37i3.3240>

Construct validity and internal consistency of three factor structures and two scoring methods of the 12-item General Health Questionnaire

Introduction: The analysis of the psychometric properties of an instrument is important for the control of measurement bias.

Objective: To compare the construct validity and internal consistency of three factorial structures of the General Health Questionnaire-12.

Materials and methods: We conducted a validation study in a nested sample of 483 dentistry students from Cartagena, Colombia. The instrument was applied along with questions about personal factors. The construct validity of the following structures was evaluated: one-dimensional, one-dimensional with correction of response bias, and the two and three-factor structure, and these scoring systems: Likert scale (0-1-2-3) and dichotomous (0-0-1-1) scoring method. Validity was determined by confirmatory factorial analysis, and the internal consistency with Cronbach's alpha coefficient for the Likert scale and the Kuder-Richardson Formula 20 for the dichotomous scoring using the Stata™, v. 13.2 (Stata Corp, USA) and Mplus™, v.7.11 (Muthen and Muthen, USA) software.

Contribución de los autores:

Miguel Simancas-Pallares: concepción de la idea y análisis de la información

Katherine Arrieta: concepción de la idea

Todos los autores participaron en la recolección de la información y la estructuración del manuscrito.

Results: Adjustment indexes revealed that the best model was the one-dimensional one with correction of response bias based on a dichotomous scoring (degrees of freedom=36; chi square=52.432; root mean square error of approximation=0.03; 90% CI: 0.008-0.048; comparative fit index=0.982; Tucker-Lewis index=0.966). The internal consistency was 0.70.

Conclusions: The adjustment of the model in this study allowed for the control of measurement bias and guaranteed external validity of results when using the General Health Questionnaire-12 to evaluate mental conditions in this population. The psychometric properties of the scales should be critically evaluated before results analysis.

Key words: validation studies; factor analysis, statistical; bias (epidemiology); psychometrics; mental health; students, dental.

doi: <https://doi.org/10.7705/biomedica.v37i3.3240>

El cuestionario de salud general (*General Health Questionnaire*, GHQ) es un instrumento que permite la valoración de la morbilidad psicológica en la población general (1,2). Propuesto por Goldberg en 1972 para la valoración del estado mental de los profesionales de la salud, cuenta con versiones de 60, 30, 23 y 12 ítems (3).

El GHQ-12 es la versión más empleada, ya que permite una evaluación rápida de trastornos mentales comunes tan precisa como la de sus predecesores. Se compone de seis ítems redactados de forma positiva y seis redactados de forma negativa (4). Las preguntas pueden puntuarse mediante la modalidad dicotómica binaria (0-0-1-1) propuesta originalmente para el cuestionario, o mediante una escala del tipo de Likert (0-1-2-3), en la cual los ítems positivos se evalúan de esta forma pero la puntuación se invierte para los negativos (5). Para corregir este efecto, se diseñó el sistema binario corregido (0-1-1-1, para los ítems negativos y 0-0-1-1, para los positivos) (6).

Varios autores han señalado que se trata de una escala que representa un constructo multifactorial, a pesar de que los hallazgos iniciales de Goldberg la planteaban como un constructo unidimensional (1,7). Andrich, *et al.*, (8) determinaron que la escala se componía de dos dominios denominados disforia general (ansiedad y depresión) y disfunción social, y que cada dominio estaba conformado por ítems negativos y positivos, respectivamente. Posteriormente, Graetz (9) determinó que la escala se componía de tres factores: ansiedad y depresión,

disfunción social y pérdida de confianza. Algunos autores incluso han evidenciado la presencia de un cuarto factor, la anhedonia (10,11).

En el 2008, Hankins comparó diversos métodos de puntuación y las estructuras de los factores del GHQ-12 mediante un análisis factorial confirmatorio, y determinó que la forma de redacción de los ítems negativos y la ambigüedad en las opciones de respuesta en el cuestionario creaban un efecto conocido como sesgo de respuesta, lo cual podía indicar una naturaleza multidimensional espuria del constructo (12). Concluyó que correlacionar la varianza de los errores de los ítems negativos permitía evidenciar de forma correcta su estructura de factores (13). En este sentido, la mejor forma de puntuación es la propuesta originalmente (0-0-1-1), con lo cual el constructo se evalúa como unidimensional.

Se han hecho múltiples esfuerzos para replicar las estructuras factoriales propuestas por Goldberg, *et al.*, Andrich, *et al.*, y Graetz (14,15). Sin embargo, la inestabilidad del constructo, las características de la población y las particularidades de la redacción de sus ítems, no permiten conclusiones definitivas sobre los múltiples factores encontrados. Además, en diversos estudios se ha empleado de forma indiscriminada el GHQ-12 con una escala de Likert para la puntuación (16), pero muy pocos de estos estudios se han hecho con el cuestionario en español y utilizando un análisis factorial confirmatorio para determinar su validez, por lo cual la capacidad de determinar la validez real del constructo en la población evaluada y la posibilidad de generalizar los resultados encontrados es limitada (17). En Colombia se han hecho estudios exploratorios sobre la estructura de factores del GHQ-12, pero ninguno ha recurrido al análisis factorial confirmatorio, razón por la cual no existe evidencia científica en el país sobre la validez de esta escala (18).

En este sentido, el objetivo del presente estudio fue comparar la validez de constructo y la consistencia interna de tres estructuras factoriales y dos

Correspondencia:

Miguel Simancas-Pallares, Departamento de Investigación, Facultad de Odontología, Universidad de Cartagena, Avenida del Consulado, Calle 30 N° 48-152, Campus Ciencias de la Salud, Zaragocilla, piso 2, oficina 301, Cartagena de Indias, Colombia
Teléfono: (575) 669 8172, extensión 110
msimancasp@unicartagena.edu.co

Recibido: 09/03/16; aceptado: 08/10/16

métodos de puntuación (del tipo de la escala de Likert y dicotómica binaria) del cuestionario de salud general de 12 ítems.

Materiales y métodos

Se hizo un estudio de validación en una muestra anidada de 483 estudiantes de odontología de dos universidades de Cartagena, Colombia. La investigación se desarrolló en el marco de un estudio multicéntrico desarrollado en tres facultades de Odontología de Cartagena, para determinar la presencia de trastornos mentales comunes y los factores relacionados con su aparición (19). Debido a las similitudes sociodemográficas, personales y académicas de los participantes de cada centro, se seleccionó una muestra aleatoria conformada por poco más de la mitad (56,16 %) de los estudiantes.

Los estudiantes respondieron de forma anónima la encuesta que, además del GHQ-12, contenía preguntas que indagaban por la presencia de factores personales y académicos de interés. Se incluyeron estudiantes matriculados, activos académicamente y que voluntariamente desearon participar en la investigación.

El GHQ-12 consta de 12 ítems redactados en forma de pregunta que indagan sobre la presencia de síntomas emocionales (de ansiedad y depresión) en las cuatro semanas anteriores. Seis de estas preguntas están redactadas de forma positiva (ítems 1, 2, 7, 10, 11 y 12) y, seis, de forma negativa (3, 4, 5, 6, 8 y 9); las opciones de respuesta fueron las siguientes: nunca, a veces, muchas veces y siempre. Las preguntas del cuestionario se enumeran en el cuadro 1. La versión en español empleada en el estudio se tomó de una investigación previamente realizada en Colombia con este instrumento (20).

Para la evaluación de los síntomas, se pueden emplear diversos métodos de puntuación: uno del tipo de la escala de Likert (ítems positivos: 0-1-2-3; ítems negativos: 3-2-1-0), un método dicotómico (ítems positivos: 0-0-1-1; ítems negativos: 1-1-0-0) o un método dicotómico corregido (ítems negativos: 0-1-1-1; ítems positivos: 0-0-1-1). Si se emplea la forma de puntuación del tipo de la escala de Likert, los puntajes varían de 0 a 3; en este sentido, el máximo puntaje posible es 36. Si se emplea una puntuación dicotómica, los puntajes pueden variar de 0 a 12.

Mediante el análisis factorial confirmatorio, se determinó la validez de constructo de las siguientes estructuras de factores: la unidimensional, la unidimensional con corrección del sesgo de respuesta

(correlación del error de la covarianza de los ítems negativos), y la de dos y tres factores. Cada una de estas estructuras se combinó con el método de puntuación dicotómico y el de tipo de la escala de Likert. De esta manera, se determinó la validez de constructo de ocho modelos.

Mediante el método de estimación de mínimos cuadrados parciales con media y varianza ajustada, se obtuvieron los índices de ajuste para cada uno de estos modelos: χ^2 al cuadrado y grados de libertad (gl), p , raíz cuadrada del error medio de aproximación y su intervalo de confianza de 90 %, índice comparativo de ajuste e índice de Tucker-Lewis. Empleando los criterios propuestos por Hu, *et al.*, se determinó que el ajuste del modelo era aceptable si presentaba los siguientes valores: p mayor de 0,05; raíz cuadrada del error medio de aproximación de 0,06 o menor; índice comparativo de ajuste e índice de Tucker-Lewis de 0,95 o mayor (21).

La consistencia interna se determinó con el coeficiente alfa de Cronbach para la puntuación del tipo de la escala de Likert y con la fórmula 20 de Kuder-Richardson para la forma dicotómica, y se calificó como aceptable cuando era de 0,70 o mayor (22). El análisis descriptivo y la estimación de la consistencia interna se hicieron con el programa Stata™, v. 13.2 para Windows (Stata Corp., College Station., TX., USA) y, el análisis factorial confirmatorio, con el Mplus™, v. 7.11 para Windows (Muthen y Muthen, Los Ángeles, CA, USA).

Cuadro 1. Nombre completo y abreviado de los ítems que componen el GHQ-12

Nº	Ítem	Nombre corto
1	¿Ha perdido el sueño por sus preocupaciones?	Pérdida de sueño
2	¿Se ha sentido constantemente tensionado?	Tensión
3	¿Se ha podido concentrar en lo que está haciendo?	Concentración
4	¿Siente que usted representa un papel útil en lo que está haciendo?	Papel útil
5	¿Ha sido capaz de afrontar sus problemas?	Afrontar problemas
6	¿Ha sido capaz de tomar decisiones?	Tomar decisiones
7	¿Ha sentido que no puede vencer sus dificultades?	Vencer dificultades
8	¿Teniendo en cuenta todo, se ha sentido razonablemente feliz?	Ser feliz
9	¿Ha sido capaz de disfrutar sus actividades diarias?	Disfrutar actividades
10	¿Se ha sentido triste o deprimido?	Sentirse triste
11	¿Ha perdido la confianza en usted mismo?	Perder la confianza
12	¿Ha estado pensando que usted no vale nada?	No valer nada

Consideraciones éticas

Este estudio contó con la aprobación del Comité de Ética en Investigación de la Universidad de Cartagena. En cumplimiento de las normas nacionales e internacionales para investigación en salud, todos los participantes firmaron un consentimiento informado después de recibir información suficiente y entender los objetivos del estudio.

Resultados

La mayoría de los estudiantes incluidos en la muestra final estaba en los semestres del ciclo básico (53,6 %), más de la mitad eran mujeres (66,2 %) y la edad promedio fue de 20,8 años (desviación estándar, DE=3,01) (cuadro 2).

Validez de constructo

Según los resultados, el mejor ajuste se obtuvo con la estructura unidimensional con corrección del sesgo de respuesta y errores de la covarianza correlacionados para los ítems negativos y la forma binaria de puntuación ($gl=36$; ji al cuadrado=52,432; $p<0,0001$; raíz cuadrada del error medio de aproximación=0,031; $IC_{90\%}$ 0,127-0,148; índice comparativo de ajuste=0,982; índice de Tucker-Lewis=0,966). Los índices de ajuste obtenidos para cada uno de los ocho modelos evaluados se presentan en el cuadro 3 y en la figura 1 se muestra el diagrama de vías para el mejor ajuste.

Consistencia interna

Para la forma de puntuación del tipo de la escala de Likert, la consistencia interna medida con el alfa de Cronbach fue de 0,7415 ($IC_{95\%}$ 0,712-0,771), en

tanto que, para la puntuación dicotómica obtenida con la fórmula 20 de Kuder-Richardson, fue de 0,70 ($IC_{95\%}$ 0,667-0,735).

Discusión

El cuestionario de salud general de 12 ítems (GHQ-12) se diseñó como una forma fácil, rápida y económica de evaluar el estado emocional durante las cuatro semanas previas en pacientes o usuarios de servicios médicos, y se utiliza en estudios epidemiológicos (23-25). En investigaciones anteriores, se ha sugerido la existencia de estructuras multidimensionales en el GHQ-12, aunque con datos contradictorios sobre la validez de cada uno de los modelos (26). Los resultados del presente estudio responderían a la interacción de tres fuentes principales de sesgos: categorías de respuesta ambigua para los ítems negativos, múltiples sistemas o formas de puntuación y métodos de estimación inapropiados (27).

Aunque la mayoría de los estudios en psicometría han centrado sus esfuerzos en el análisis factorial exploratorio para hacer inferencias sobre la validez de los modelos derivados de algún constructo, no deben descartarse las técnicas estadísticas basadas en el modelado de ecuaciones estructurales, o análisis factorial confirmatorio, para la correcta determinación de la validez de constructo. En este sentido, la importancia del análisis factorial exploratorio radica en que de él se derivan las hipótesis que serán puestas a prueba en el análisis confirmatorio (28).

En el presente estudio, dicho análisis se empleó para comparar la validez de constructo de tres estructuras de factores del GHQ-12 y dos sistemas de puntuación mediante el empleo de métodos de estimación apropiados para el tipo de variable evaluada (29). En varios de los estudios de comparación de las estructuras factoriales y los métodos de puntuación del GHQ-12, se ha evaluado el método de puntuación sugerido por Goodchild, *et al.* (0-1-1-1), conocido como GHQ-correctado; no obstante, se ha comprobado que el sesgo de respuesta persistía, por lo cual no se incluyó en la presente investigación. En este método, la puntuación 0-0-1-1 se aplica a los ítems redactados de forma positiva, pero los ítems negativos se puntúan 0-1-1-1, con lo cual se colapsan las categorías 2, 3 y 4. Con el método de puntuación GHQ-correctado debe prestarse atención a las respuestas que a veces se otorgan a un ítem negativo, pues podrían indicar la presencia de un problema en lugar de ser señal de buena salud (14).

Cuadro 2. Características sociodemográficas de la muestra estudiada

	n	%
Sexo		
Femenino	320	66,2
Masculino	163	33,8
Estado civil		
Soltero	459	95
Otro	24	5
Ciclo académico		
Básico	259	53,6
Clínico	224	46,4
Actividad		
Solo estudia	450	93,7
Trabaja y estudia	33	6,8
Nivel socioeconómico		
Alto	4	0,8
Medio	92	19
Bajo	387	80,1

Cuadro 3. Índices de ajuste obtenidos en cada uno de los modelos evaluados mediante análisis factorial complementario

	Unidimensional		Unidimensional'		Andrich, <i>et al.</i> (8)		Graetz (9)	
	Likert	0-0-1-1	Likert	0-0-1-1	Likert	0-0-1-1	Likert	0-0-1-1
Grados de libertad	54	54	36	36	53	53	51	51
Ji al cuadrado	577,599 [‡]	261,678 [‡]	63,799 [‡]	52,432 [‡]	534,95 [‡]	243,353 [‡]	516,232 [‡]	235,91 [‡]
RCEMA	0,142	0,089	0,040	0,031	0,137	0,086	0,137	0,087
IC _{90%}	0,131-0,152	0,079-0,100	0,023-0,056	0,008-0,048	0,127-0,148	0,075-0,097	0,127-0,148	0,076-0,098
Índice comparativo de ajuste	0,700	0,768	0,984	0,982	0,724	0,787	0,733	0,793
Índice de Tucker-Lewis	0,633	0,716	0,971	0,966	0,656	0,735	0,655	0,733

* Modelo con corrección del sesgo de respuesta y errores correlacionados de los ítems negativos; ‡: modelo estadísticamente significativo: p<0,0001
RCEMA: raíz cuadrada del error medio de aproximación; IC_{90%}: intervalo de confianza de 90% para la RCEMA

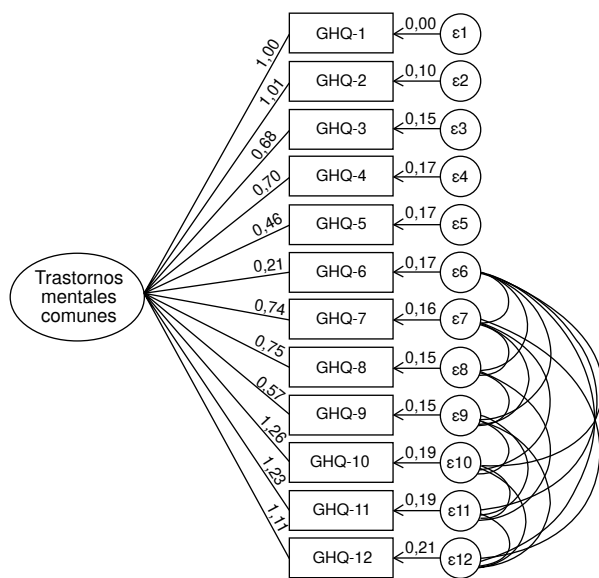


Figura 1. Diagrama de vías del análisis factorial confirmatorio de la estructura unidimensional con corrección del sesgo de respuesta (correlación del error de la covarianza para ítems negativos) y opción de respuesta dicotómica binaria del cuestionario de salud general de 12 ítems en español. La letra griega epsilon en los círculos indica los errores asociados a los estimadores derivados del modelo confirmatorio de análisis factorial (cuadros con los nombres abreviados de los ítems), y el constructo se representa con un óvalo central a la derecha. Los valores de los estimadores aparecen sobre las líneas continuas entre el constructo y cada uno de los ítems, y los valores de error de cada estimador se indican con las flechas entre los ítems y los errores.

Una de las justificaciones para emplear la puntuación del tipo de la escala de Likert es que produce puntajes con una aproximación a la distribución normal y ello permite hacer análisis estadísticos paramétricos (30). Sin embargo, las puntuaciones del cuestionario del tipo de la escala de Likert y del GHQ-correctado se ven afectadas por el sesgo de respuesta, lo cual se evidencia porque, cuando

se emplea el método de puntuación dicotómico original, los índices de ajuste derivados del modelo unidimensional son adecuados, como ocurrió con los obtenidos en el presente estudio (gl=36; ji al cuadrado=52,432; raíz cuadrada del error medio de aproximación=0,03; IC_{90%}=0,008-0,048; índice comparativo de ajuste=0,982; índice de Tucker-Lewis=0,966).

Estos resultados son congruentes con lo sugerido por Hankins (12,13) en el sentido de que los puntajes de los ítems redactados de forma negativa contienen un sesgo de respuesta evidenciado por la alta varianza encontrada en ellos. En general, este sesgo se elimina cuando se emplea el método de puntuación original con base en un enfoque conocido como método de efectos o corrección del sesgo de respuesta (correlación de la covarianza de los errores de los ítems negativos). Los resultados obtenidos sugieren que el sesgo de respuesta puede introducir un grado de error de medición inaceptable en la psicometría clínica, el cual no había sido previamente reconocido.

A la luz de los resultados obtenidos en el presente estudio, el ajuste aceptable obtenido en otras investigaciones para los modelos multidimensionales con el sistema de puntuación GHQ-correctado (31), se debe al agrupamiento artificial de los ítems redactados de forma positiva frente a los redactados de forma negativa, situación conocida como error de reificación (13,30). Así pues, en futuros estudios sobre la estructura de factores de este instrumento deben tenerse en cuenta los hallazgos del presente estudio, con el fin de evitar conclusiones inapropiadas sobre su dimensión (30).

La separación dimensional entre los ítems positivos y los negativos puede explicarse también por las diferencias en los intervalos entre las categorías

de respuesta. Las etiquetas para las opciones de respuesta de los ítems positivos parecen ser bipolares y oscilar entre un polo del atributo (“casi siempre”) y el opuesto (“a veces”), mientras que las opciones de respuesta para los ítems negativos parecen ser unipolares y oscilar entre la ausencia del atributo (“nunca”) y su presencia (“siempre”). Para los ítems negativos, el formato unipolar parece presentar la opción lógica (por ejemplo, “a veces” para un ítem negativo como “perder la confianza” podría implicar una doble negación que puede confundir al sujeto) (27). Varios autores han sugerido que los formatos de respuesta bipolar pueden causar problemas cuando se trata de hacer mediciones negativas de los constructos (32).

Por otro lado, la consistencia interna fue aceptable para las dos formas de puntuación evaluadas en el presente estudio. Los estimadores convencionales de confiabilidad, como el coeficiente alfa de Cronbach, pueden sobrestimar o subestimar la confiabilidad si los supuestos de la teoría clásica no se cumplen. Dichos supuestos, entonces, deben evaluarse y, en caso de ser necesario, debe emplearse un método alternativo de estimación de la confiabilidad del instrumento.

Los resultados del presente estudio revelaron una sobreestimación de la confiabilidad del GHQ-12 al emplear la forma de puntuación del tipo de la escala de Likert (alfa=0,74), dado que el estimador se calculó sin tener en cuenta la correlación de la covarianza de los errores de los ítems redactados de forma negativa (12). En estudios previos en Bucaramanga, se obtuvieron valores de confiabilidad del instrumento para la forma de puntuación del tipo de la escala de Likert (alfa=0,779) y para la binaria (alfa=0,70) similares a los encontrados en el presente estudio (alfa=0,74 y alfa=0,70, respectivamente) (20). En otro estudio en Medellín en pacientes hospitalizados, los valores del alfa de Cronbach también fueron similares (alfa=0,78) a los obtenidos en el presente estudio (18).

Los resultados de este estudio coinciden con los de otras investigaciones y sustentan el carácter unidimensional del constructo siempre que se aplique la corrección del sesgo de respuesta para los ítems redactados de forma negativa. De todas maneras, es evidente la necesidad de llevar a cabo nuevos estudios para probar la estructura de factores del GHQ-12. Es imperativo que este tipo de enfoques metodológicos se utilice antes del análisis de resultados en los estudios epidemiológicos en los que se emplee el instrumento.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Financiación

Ninguna.

Referencias

1. **Goldberg DP.** The detection of psychiatric illness by questionnaire: A technique for the identification and assessment of non-psychotic psychiatric illness. London: Oxford University Press; 1972. p. 156.
2. **Namjoo S, Shagaghi A, Sarbaksh P, Allahverdipour H, Pakpour AH.** Psychometric properties of the General Health Questionnaire (GHQ-12) to be applied for the Iranian elder population. *Aging Ment Health.* 2016;1-5. <https://doi.org/10.1080/13607863.2016.1196337>
3. **Golderberg D, Williams P.** A user's guide to the General Health Questionnaire. Windsor, UK: NFER-Nelson; 1988.
4. **Rocha KB, Pérez K, Sanz MR, Borrell C, Llandrich JO.** Propiedades psicométricas y valores normativos del *General Health Questionnaire* (GHQ-12) en población general española. *Int J Clin Health Psychol.* 2011;11:125-39.
5. **Petkovska MS, Bojadziev MI, Stefanovska VV.** Reliability, validity and factor structure of the 12-item General Health Questionnaire among general population. *Open Access Maced J Med Sci.* 2015;3:478-83. <https://doi.org/10.3889/oamjms.2015.075>
6. **Goodchild M, Duncan-Jones P.** Chronicity and the General Health Questionnaire. *Br J Psychiatry.* 1985;146:55-61.
7. **Rahmati-Najarkolaei F, Raiisi F, Rahnama P, Gholami-Fesharaki M, Zamani O, Jafari MR, et al.** Factor structure of the Iranian version of 12-item general health questionnaire. *Iran Red Crescent Med J.* 2014;16:e11794. <https://doi.org/10.5812/ircmj.11794>
8. **Andrich D, van Schoubroeck L.** The General Health Questionnaire: A psychometric analysis using latent trait theory. *Psychol Med.* 1989;19:469-85.
9. **Graetz B.** Multidimensional properties of the General Health Questionnaire. *Soc Psychiatry Psychiatr Epidemiol.* 1991;26:132-8
10. **Smith AB, Fallowfield LJ, Stark DP, Velikova G, Jenkins V.** A Rasch and confirmatory factor analysis of the general health questionnaire (GHQ)--12. *Health Qual Life Outcomes.* 2010;8:45. <https://doi.org/10.1186/1477-7525-8-45>
11. **van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH.** Validation of online psychometric instruments for common mental health disorders: A systematic review. *BMC Psychiatry.* 2016;16:45. <https://doi.org/10.1186/s12888-016-0735-7>
12. **Hankins M.** The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health.* 2008;8:355. <https://doi.org/10.1186/1471-2458-8-355>
13. **Hankins M.** The factor structure of the twelve item General Health Questionnaire (GHQ-12): The result of negative phrasing? *Clin Pract Epidemiol Ment Health.* 2008;4:10. <https://doi.org/10.1186/1745-0179-4-10>

14. **Gelaye B, Tadesse MG, Lohsoonthorn V, Lertmeharit S, Pensuksan WC, Sánchez SE, et al.** Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *J Affect Disord.* 2015;187:197-202. <https://doi.org/10.1016/j.jad.2015.08.045>
15. **Padrón A, Galán I, Durban M, Gandarillas A, Rodríguez-Artalejo F.** Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Qual Life Res.* 2012;21:1291-8. <https://doi.org/10.1007/s11136-011-0038-x>
16. **Abu-Ghazaleh SB, Rajab LD, Sonbol HN.** Psychological stress among dental students at the University of Jordan. *J Dent Educ.* 2011;75:1107-14.
17. **Brabete AC.** The 12-Item General Health Questionnaire (GHQ-12): Translation and adaptation study of the Romanian version. *RIDEP.* 2014;37:11-29.
18. **Villa G, Cristina I, Zuluaga-Arboleda C, Restrepo-Roldán LF.** Propiedades psicométricas del Cuestionario de Salud General de Goldberg, GHQ-12, en una institución hospitalaria de la ciudad de Medellín. *Avances en Psicología Latinoamericana.* 2013;31:532-45.
19. **Simancas MA, Arrieta KM.** Common mental disorders and related factors in undergraduate and graduate students from three dental faculties in Cartagena, Colombia. *Study protocol. Journal of Oral Research.* 2014;3:178-83. <https://doi.org/10.17126/JORALRES.2014.042>
20. **Campo-Arias A.** Cuestionario general de salud-12: análisis de factores en población general de Bucaramanga, Colombia. *Iatreia.* 2007;20:29-36.
21. **Hu LT, Bentler PM.** Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6:1-55. <https://doi.org/10.1080/10705519909540118>
22. **Tavakol M, Dennick R.** Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53-5. <https://doi.org/10.5116/ijme.4dfb.8dfd>
23. **Hassan BK, Werneck GL, Hasselmann MH.** Maternal mental health and nutritional status of six-month-old infants. *Rev Saúde Pública.* 2016;50:7. <https://doi.org/10.1590/S1518-8787.2016050006237>
24. **Rutagarama O, Gelaye B, Tadesse MG, Lemma S, Berhane Y, Williams MA.** Risk of common mental disorders in relation to symptoms of obstructive sleep apnea syndrome among Ethiopian college students. *J Sleep Disord Treat Care.* 2015;4. <https://doi.org/10.4172/2325-9639.1000161>
25. **Lopes CS, Abreu G de A, dos Santos DF, Menezes PR, de Carvalho KM, Cunha Cde F, et al.** ERICA: Prevalence of common mental disorders in Brazilian adolescents. *Rev Saúde Pública.* 2016;50(Suppl.1):14s. <https://doi.org/10.1590/S01518-8787.2016050006690>
26. **Aguado J, Campbell A, Ascaso C, Navarro P, García-Esteve L, Luciano JV.** Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment.* 2012;19:517-25. <https://doi.org/10.1177/1073191110388146>
27. **Rey JJ, Abad FJ, Barrada JR, Garrido LE, Ponsoda V.** The impact of ambiguous response categories on the factor structure of the GHQ-12. *Psychol Assess.* 2014;26:1021-30. <https://doi.org/10.1037/a0036468>
28. **Williams B, Onsmann A, Brown T.** Exploratory factor analysis: A five-step guide for novices. *Australian Journal of Paramedicine.* 2010;8:1-13.
29. **Brown TA.** Confirmatory factor analysis for applied research. New York: The Guilford Publications; 2015.
30. **Molina JG, Rodrigo MF, Losilla JM, Vives J.** Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychol Assess.* 2014;26:1031-7. <https://doi.org/10.1037/a0036472>
31. **Urzúa A, Caqueo-Úrizar A, Bargsted M, Irrázaval M.** ¿Afecta la forma de puntuación la estructura factorial del GHQ-12? Estudio exploratorio en estudiantes iberoamericanos. *Cad Saúde Pública.* 2015;31:1305-12. <https://doi.org/10.1590/0102-311X00122913>
32. **Mazaheri M, Theuns P.** Effects of varying response formats on self-ratings of life-satisfaction. *Soc Indic Res.* 2009;90:381-95. <https://doi.org/10.1007/s11205-008-9263-2>