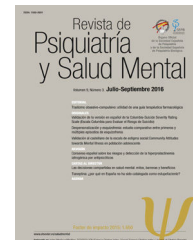




Revista de Psiquiatría y Salud Mental

www.elsevier.es/saludmental



ORIGINAL

Estudio comparativo de los formatos en lápiz y papel y electrónicos de los cuestionarios GHQ-12, WHO-5 y PHQ-9



María Luisa Barrigón^a, Ana María Rico-Romano^a, Marta Ruiz-Gomez^b, David Delgado-Gomez^c, Igor Barahona^d, Fuensanta Aroca^d, Enrique Baca-García^{a,b,e,f,g,h,i,*} y MEmind Study Group^{abhijk}◇

^a Departamento de Psiquiatría, Hospital Universitario Fundación Jiménez Díaz, IIS-Fundación Jiménez Díaz, Madrid, España

^b Departamento de Psiquiatría, Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, España

^c Departamento de Estadística, Universidad Carlos III, Getafe, Madrid, España

^d Instituto de Matemáticas, Universidad Nacional Autónoma de México, México

^e CIBERSAM, Madrid, España

^f Universidad Autónoma de Madrid, Madrid, España

^g Psychiatry Department, Columbia University, Nueva York, Estados Unidos

^h Departamento de Psiquiatría, Hospital Universitario Infanta Elena, Valdemoro, Madrid, España

ⁱ Departamento de Psiquiatría, Hospital General de Villalba, Collado Villalba, Madrid, España

^j Hospital 12 de Octubre, Madrid, España

^k AGC Salud Mental, Área Sanitaria 3, Avilés, Asturias, España

Recibido el 26 de julio de 2016; aceptado el 20 de diciembre de 2016

Disponible en Internet el 22 de febrero de 2017

PALABRAS CLAVE

GHQ-12;
WHO-5;
PHQ-9;
Psicometría;
Fiabilidad

Resumen

Introducción: El auge de la telemedicina en el campo de la salud mental está haciendo que el uso de instrumentos psicométricos, tradicionalmente basados en un soporte de «lápiz-y-papel», se adapte al formato electrónico. El objetivo de este trabajo es verificar si los 2 formatos de conocidos instrumentos como las escalas GHQ-12, WHO-5 y PHQ-9 son intercambiables.

Material y métodos: Cuarenta y siete voluntarios completaron los cuestionarios GHQ-12, WHO-5 y PHQ-9 en papel y en las siguientes 24h en su versión electrónica a través de la web www.memind.net (24 en formato electrónico-Likert y 23 en formato electrónico-slider). Se midió la consistencia interna mediante el índice α -Cronbach y el coeficiente omega, la fiabilidad test-retest mediante el coeficiente de correlación intraclase (CCI), el grado de acuerdo de los ítems mediante el coeficiente kappa ponderado y comparamos la estructura dimensional de los 2 formatos con el índice de ajuste comparativo (IAC).

Resultados: La consistencia interna de los instrumentos fue mayor de 0,8 para todos los instrumentos a excepción del PHQ-9. Los CCI oscilaron entre 0,655 para PHQ-9

* Autor para correspondencia.

Correo electrónico: ebacgar2@yahoo.es (E. Baca-García).

◇ Más información sobre los colaboradores del MEmind Study Group está disponible en el [anexo](#).

<http://dx.doi.org/10.1016/j.rpsm.2016.12.002>

1888-9891/© 2017 SEP y SEPB. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

lápiz-y-papel/electrónico-slider y 0,901 para GHQ-12 lápiz-y-papel/electrónico-slider. El acuerdo entre los ítems en las versiones en lápiz-y-papel y electrónica-Likert fue variable, desde acuerdos muy bajos en el del ítem PHQ-1 (κ ponderada = 0,143; $p = 0,384$) hasta acuerdos altos en el ítem PHQ-5 (κ ponderada = 0,769; $p = 0,000$). La equivalencia arrojada con el IAC fue adecuada.

Conclusiones: Con la excepción del PHQ-9 electrónico-Likert, los cuestionarios mantienen su estructura en la versión electrónica. Encontramos discrepancias en el acuerdo entre los ítems, lo que apoya la idea de que el paso de un instrumento diseñado en papel al medio digital no es un proceso automático, sino que requiere un proceso de adaptación y de verificación sobre el tipo de diseño en formato electrónico.

© 2017 SEP y SEP. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

GHQ-12;
WHO-5;
PHQ-9;
Psychometrics;
Reliability

Comparative study of pencil-and-paper and electronic formats of GHQ-12, WHO-5 and PHQ-9 questionnaires

Abstract

Introduction: The increase in telemedicine in the mental health field has led to psychometric instruments changing from paper-and-pencil administration to an electronic format. A study is performed to determine if both formats are equivalent for well-known questionnaires such as GHQ-12, WHO-5, and PHQ-9.

Material and methods: Forty-seven volunteers completed GHQ-12, WHO-5 and PHQ-9 questionnaires in paper-and-pencil format, and in the following 24h they completed their electronic versions via the web site www.memind.net. An electronic-Likert format was used by 24 participants, and 23 used an electronic-slider format. Internal consistency was measured by α -Cronbach index and omega coefficient, and test-retest was measured by the intraclass correlation coefficient (ICC). Agreement between individual items was compared using Weighted Kappa coefficients, and dimensional structure between formats using the Comparative Fit Index (CFI).

Results: Internal consistency was higher than 0.8 for GHQ-12 and WHO-5. The ICC ranged between 0.655 for PHQ-9 paper-and-pencil/electronic-slider and 0.901 for GHQ-12 paper-and-pencil/electronic-slider. Agreement for individual items in paper-and-pencil and electronic-Likert versions was variable, ranging from low agreement in PHQ-1 (weighted $\kappa = 0.143$; $P = .384$) to high agreement in PHQ-5 (weighted $\kappa = 0.769$; $P = .000$). The CFI results showed an adequate equivalence between formats.

Conclusions: Except for the PHQ-9 electronic-Likert, questionnaires keep their structure in electronic formats. Discrepancies were found in items agreement. This study supports previous works indicating that the change from paper-and-pencil to electronic formats is not an immediate process, and needs a proper adaptation.

© 2017 SEP y SEP. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

En los últimos años las tecnologías de la telecomunicación y computación se han ido incorporando al campo de la salud, dando lugar a una nueva disciplina que se ha denominado telemedicina¹. La telemedicina también ha alcanzado la salud mental y actualmente es un área en expansión², siendo un campo de especial interés la evaluación de síntomas por medio de herramientas electrónicas como ordenadores, smartphones o wearables³, de modo que lo que clásicamente se ha estudiado en el campo de la psicometría con «lápiz y papel» está pasando a hacerse en formato electrónico. Distintos estudios han hallado que los participantes se encuentran más cómodos ante los cuestionarios electrónicos, los responden con mayor facilidad y se

pierden menos datos en comparación con los cuestionarios en papel⁴⁻⁷.

Cuando los instrumentos psicométricos se emplean en poblaciones con características diferentes para las que habían sido creados tiene que haber un proceso de validación y adaptación⁸, y de igual modo, en la migración del formato en lápiz y papel al formato electrónico hay iniciativas que abogan por demostrar la equivalencia entre ambos formatos en lugar de asumir una equivalencia directa⁹.

En el momento actual, la literatura existente apoya la idea de que los cuestionarios en lápiz y papel y los electrónicos arrojan datos equivalentes, tanto en medidas de resultado reportadas por los pacientes en ensayos clínicos^{4,10} como en instrumentos psicométricos. Así, 3 revisiones estudian la fiabilidad entre las versiones en lápiz y

papel y electrónica de unos 50 instrumentos psicométricos —fundamentalmente en ansiedad y depresión— encontrando una buena fiabilidad en general, aunque con discrepancias en algunos cuestionarios¹¹⁻¹³.

Cuando los cuestionarios psicométricos miden el grado de acuerdo con una frase es frecuente que las respuestas se presenten en una escala tipo Likert, ofreciendo un número determinado de opciones de mayor a menor acuerdo. Una alternativa a estos cuestionarios son las escalas analógicas visuales, descritas por Hayes y Patterson ya en 1921¹⁴, en las que el encuestado señala su grado de acuerdo con lo preguntado en un gráfico lineal. Las características del formato electrónico facilitan la introducción del formato analógico visual a modo de una línea en la que el encuestado puede situarse en el punto que más le representa deslizando el ratón (de ahí la denominación «slider»). Se ha demostrado que las escalas analógicas visuales o tipo slider son una buena alternativa en los cuestionarios electrónicos/online^{15,16} y permiten una administración más sencilla y rápida y una mejor comprensión por parte de las personas que las cumplimentan¹⁷.

Un área específica de la psicometría es la del screening de trastornos mentales con cuestionarios autoadministrados. En la era de la telemedicina disponer de versiones electrónicas fiables de este tipo de instrumentos facilitaría el acceso a ellos de una población cada vez más familiarizada con la tecnología. Esto es especialmente importante si tenemos en cuenta que la población de *nativos digitales*¹⁸ empieza a ser usuaria de los servicios de salud, donde los profesionales que los reciben siguen siendo en su mayoría *inmigrantes digitales*¹⁸.

Nuestro objetivo en este trabajo es analizar si los cuestionarios de screening Cuestionario de Salud General (GHQ-12), Índice de Bienestar de la Organización Mundial de la Salud (WHO-5) y Cuestionario de Salud del Paciente (PHQ-9) mantienen las mismas propiedades psicométricas en el paso de la versión en lápiz y papel a la electrónica tipo Likert (electrónica-Likert) y electrónica tipo slider (electrónica-slider). Nuestra hipótesis es que las versiones en lápiz y papel y electrónica de cada una de estas escalas serán intercambiables.

Material y métodos

Participantes y procedimiento

La muestra consistió en 47 voluntarios no clínicos de edades comprendidas entre los 18 y 24 años, y mayoritariamente mujeres (35 del total), divididos en 2 grupos de 24 y 23 participantes. Todos ellos eran estudiantes de Grado de Enfermería de la Universidad Autónoma de Madrid.

En el mes de mayo de 2016 los participantes respondieron a una entrevista autoaplicada que incluía las versiones en papel de los siguientes instrumentos de screening: GHQ-12, WHO-5 y PHQ-9 en sus correspondientes aulas al final de la última clase de la mañana. Todos los estudiantes accedieron a participar. Los cuestionarios fueron completados de forma anónima, sin recibir ningún tipo de compensación por ello. Tras devolver los cuestionarios rellenos al investigador se les proporcionaron claves de acceso para la herramienta multiplataforma web

www.memind.net, donde cumplimentaron los mismos cuestionarios en las siguientes 24 h en sus versiones electrónicas; 24 participantes rellenan una versión electrónica-Likert y 23 una electrónica-slider.

El estudio fue aprobado por el comité de ética para la investigación clínica del Hospital Universitario Fundación Jiménez Díaz. Tras ser informados sobre la naturaleza del estudio, todos los participantes dieron su consentimiento informado por escrito previamente a su participación. Los resultados han sido presentados siguiendo la declaración de transparencia propuesta por Catalá-López et al.^{19,20}.

Instrumentos

Los cuestionarios empleados fueron la versión de 12 ítems del GHQ-12²¹, el WHO-5²² y el PHQ-9 de screening de depresión²³.

En el momento inicial los participantes rellenan las versiones originales en lápiz y papel y en un segundo tiempo las versiones electrónicas —de desarrollo propio— electrónica-Likert y electrónica-slider.

Cuestionario de Salud General (GHQ-12)

Es un cuestionario autoadministrado, diseñado para su uso en medios clínicos para detectar individuos con posibles trastornos psiquiátricos. Lo conforman 12 ítems, 6 formulados en positivo y 6 formulados en negativo, que se contestan en una escala tipo Likert de 4 puntos, con las respuestas —puntuadas de 0 a 3—: «más que lo habitual», «igual que lo habitual», «menos que lo habitual» y «mucho menos que lo habitual» para las preguntas positivas y «no, en absoluto», «no más que lo habitual», «algo más que lo habitual» y «mucho más que lo habitual» para los ítems negativos. Puntuaciones más altas indican peor salud. Se han desarrollado distintos métodos de puntuación: el método estándar (GHQ-0011), la puntuación corregida (GHQ-0111) y la puntuación Likert (GHQ-0123).

El método estándar es el que se considera más apropiado para la identificación de casos; con él se obtiene una puntuación de entre 0 y 12, situándose el punto de corte entre 2 y 3²⁴; este fue el método usado en la comparación entre el formato en lápiz y papel y el electrónico-Likert. Para el grupo de participantes en quienes se comparó la versión en lápiz y papel y la versión electrónica-slider, se usó la puntuación GHQ-0123.

El GHQ-12 se ha validado en 15 países y ha presentado una buena fiabilidad, con alfas de Cronbach de entre 0,82 y 0,86²⁵. En población española, distintos estudios han encontrado una buena fiabilidad en población general con alfas de Cronbach de 0,76²⁶ y 0,86²⁵.

Índice de Bienestar de la OMS (WHO-5)

Es una escala autoadministrada de 5 ítems, usada para proporcionar una medida de la sensación de bienestar. Las 5 frases que la componen, con la particularidad de estar formuladas en positivo, son las siguientes: (1) «Me he sentido alegre y de buen humor», (2) «Me he sentido tranquilo y relajado», (3) «Me he sentido activo y enérgico», (4) «Me he despertado fresco y descansado» y (5) «Mi vida cotidiana ha estado llena de cosas que me interesan». Está disponible de forma gratuita en varios idiomas en

<http://www.who-5.org/>. Para cada uno de los ítems, se puntúa el grado de bienestar en las 2 semanas previas en una escala tipo Likert del 0 (nunca) al 5 (todo el tiempo); la puntuación total varía de 0 a 25, de modo que puntuaciones más altas se relacionan con mayor sensación de bienestar y puntuaciones menores de 13 se han relacionado con depresión²⁷.

La WHO-5 es una escala utilizada en todo el mundo y no solo en el campo de la salud mental, sino también en el de la salud general. Ha demostrado una validez adecuada como herramienta de screening de depresión y como medida de evolución en diversos ensayos clínicos²⁷. En concreto, en un estudio en población activa adulta europea, que incluía población española, la consistencia interna fue de 0,82 (alfa de de Cronbach)²⁸, y en población española de más de 65 años la consistencia interna fue de 0,86²⁹. En la última encuesta europea de calidad de vida disponible, de 2012, la puntuación media en España fue de 65,4³⁰.

Cuestionario sobre la Salud del Paciente-9 (PHQ-9)

Es un cuestionario autoadministrado usado para el screening de depresión, formado por 9 ítems formulados a partir de los criterios diagnósticos de depresión del DSM-IV; cada ítem se puntúa de 0 a 3 (0 = ningún día; 1 = varios días; 2 = más de la mitad de los días, y 3 = casi todos los días). Aunque hay otras formas de puntuación, cuando se usa como herramienta de screening de depresión la puntuación total de la PHQ-9 se calcula sumando la puntuación en cada ítem y oscila entre 0 y 27; puntuaciones mayores de 10 indican niveles moderados o altos de depresión²³.

Se ha demostrado que la utilidad del PHQ-9 como medida de screening de depresión es igual e incluso superior a otras medidas de depresión, obteniéndose una buena validez y fiabilidad en distintos estudios³¹. Específicamente, Kroenke et al.²³ reportaron una consistencia interna del PHQ-9 de 0,86 a 0,89. En España, Díez-Quevedo et al. validaron el PHQ-9 en 1.003 pacientes hospitalizados, demostrando un buen acuerdo entre los diagnósticos con el PHQ-9 y los diagnósticos clínicos (kappa=0,74; precisión total, 88%; sensibilidad, 87%; especificidad, 88%)³². En población de habla española destaca la validación hecha en población mexicana; en 55.000 mujeres de la cohorte *Mexican Teachers' Cohort* el cuestionario demostró una alta consistencia interna, con una alfa de Cronbach de 0,89³³.

MEmind Wellnes Tracker

Para la aplicación electrónica de las escalas se empleó la herramienta MEmind Wellness Tracker, desarrollada en el Departamento de Psiquiatría del Hospital Universitario Fundación Jiménez Díaz. Esta aplicación web está disponible en la dirección www.memind.net y funciona en todo tipo de dispositivos con acceso a Internet (ordenadores, tabletas y smartphones) con cualquier sistema operativo. La web consta de 2 interfaces, la del investigador y la del usuario. Tras entregar al investigador el cuestionario en papel, los alumnos fueron registrados en la web, se les proporcionaron las claves de acceso (usuario y contraseña), y el investigador les asignó aleatoriamente una de las 2 formas de cuestionario electrónico programado para poder completarse en las siguientes 24 h.

Tras el acceso a la plataforma y el rellenado de los cuestionarios, los datos fueron almacenados en un

servidor seguro y encriptados con Secure Socket Layer/Transport Layer Security (SSL/TLS). Solo el investigador principal (EBG) tiene acceso al servidor. MEmind usa claves encriptadas con códigos de 256 bits basadas en el algoritmo AES-256. Un auditor externo garantiza que estos niveles de seguridad cumplen los máximos requerimientos de la ley de protección de datos.

Análisis estadístico

El análisis estadístico se realizó con la versión 23.0 del paquete SPSS³⁴. En primer lugar, dado el escaso tamaño muestral y la ausencia de normalidad, se compararon las puntuaciones en cada una de las escalas entre las versiones en lápiz y papel y las electrónicas mediante la prueba de rangos con signo de Wilcoxon. A continuación, se calculó la fiabilidad utilizando índices de consistencia y de acuerdo. Se utilizó el índice alfa de Cronbach para medir la consistencia interna de los cuestionarios en los 2 formatos. Adicionalmente, y dadas las limitaciones señaladas por distintos autores del coeficiente alfa, se calculó también el coeficiente omega³⁵. Se utilizó el coeficiente de correlación intraclass (CCI) para estimar la fiabilidad test-retest; valores del CCI por debajo de 0,4 indican baja fiabilidad, entre 0,4 y 0,75 indican una fiabilidad entre regular y buena, y superiores a 0,75 indican una fiabilidad excelente³³. El grado de acuerdo entre los distintos ítems de las versiones en lápiz y papel y electrónica-Likert de los cuestionarios se midió utilizando los coeficientes kappa ponderado, y para su interpretación se utilizaron los criterios propuestos por Landis y Koch, según los cuales valores menores de 0 indican un acuerdo pobre, valores entre 0,01-0,20 un acuerdo ligero, entre 0,21-0,40 un acuerdo razonable, entre 0,41-0,60 un acuerdo moderado, entre 0,61-0,80 un acuerdo grande y entre 0,81-1,00 un acuerdo casi perfecto³⁶. Por último, con la finalidad de contrastar los modelos en papel y lápiz y electrónicos, y por consiguiente comparar su estructura dimensional, calculamos el índice de ajuste comparativo (IAC) para las 3 escalas evaluadas.

Resultados

Paso de formato en lápiz y papel a formato electrónico-Likert

No existieron diferencias entre las versiones en lápiz y papel y electrónica en ninguna de las escalas: GHQ-12 ($Z = -1,709$; $p = 0,087$); WHO-5 ($Z = -1,067$; $p = 0,286$) y PHQ-9 ($Z = -0,199$; $p = 0,842$), siendo las puntuaciones medias para cada una de ellas las siguientes: $1,96 \pm 2,66$ en lápiz y papel frente a $1,50 \pm 2,55$ en electrónica para GHQ-12; $13,63 \pm 2,66$ en lápiz y papel frente a $14,42 \pm 4,03$ en electrónica para WHO-5, y $4,87 \pm 2,80$ en lápiz y papel frente a $4,83 \pm 2,85$ en electrónica para PHQ-9.

Consistencia interna

Los coeficientes alfa de Cronbach para el GHQ-12, WHO-5 y PHQ-9 en lápiz y papel fueron de 0,872, 0,774 y 0,586, respectivamente, y para el GHQ-12, WHO-5 y PHQ-9 electrónicos fueron de 0,835, 0,835 y 0,654. Adicionalmente, los

Tabla 1 Acuerdo entre los ítems para el cuestionario GHQ-12

Ítem	Kappa ponderada	p
GHQ-1	0,478 ^a	0,001
GHQ-2	0,556 ^a	0,000
GHQ-3	0,492 ^a	0,000
GHQ-4	0,273 ^b	0,094
GHQ-5	0,400 ^b	0,001
GHQ-6	0,670 ^c	0,000
GHQ-7	0,509 ^a	0,000
GHQ-8	0,635 ^c	0,000
GHQ-9	0,563 ^a	0,000
GHQ-10	0,613 ^c	0,001
GHQ-11	0,774 ^c	0,000
GHQ-12	0,373 ^b	0,011

^a Acuerdo moderado.

^b Acuerdo razonable.

^c Acuerdo grande.

Tabla 2 Acuerdos entre los ítems para el cuestionario WHO-5

Ítem	Kappa ponderado	p
WHO-1	0,565 ^a	0,000
WHO-2	0,451 ^a	0,001
WHO-3	0,358 ^b	0,004
WHO-4	0,481 ^a	0,000
WHO-5	0,571 ^a	0,000

^a Acuerdo moderado.

^b Acuerdo razonable.

coeficientes omega para el GHQ-12, WHO-5 y PHQ-9 en lápiz y papel fueron de 0,900, 0,900 y 0,720, respectivamente, y para el GHQ-12, WHO-5 y PHQ-9 electrónicos fueron de 0,890, 0,900 y 0,730.

Fiabilidad test-retest

Los CCI entre las versiones en lápiz y papel y las electrónicas de las distintas escalas fueron de 0,802 ($p < 0,001$) para el GHQ-12, de 0,726 ($p < 0,001$) para el WHO-5 y de 0,682 ($p < 0,001$) para el PHQ-9.

Acuerdo entre los ítems individuales

En el cuestionario GHQ-12 el acuerdo entre los ítems fue como mínimo razonable, e incluso alto, para varios de ellos (tabla 1). Para el WHO-5, menos en el ítem 3 que fue razonable, el acuerdo entre ítems fue moderado (tabla 2). Finalmente, para el PHQ-9, excepto en los ítems PHQ-1 y PHQ-6, en los que el acuerdo fue bajo, el acuerdo fue al menos razonable, e incluso grande para los ítems PHQ-5 y PHQ-9 (tabla 3).

Paso de formato en lápiz y papel a formato electrónico-slider

No existieron diferencias estadísticamente significativas entre las 2 versiones al comparar las escalas WHO-5 y PHQ-9: WHO-5 ($Z = -0,974$; $p = 0,330$) y PHQ-9 ($Z = -1,601$;

Tabla 3 Acuerdos entre los ítems para el cuestionario PHQ-9

Ítem	Kappa ponderada	P
PHQ-1	0,143 ^a	0,384
PHQ-2	0,448 ^b	0,006
PHQ-3	0,516 ^b	0,000
PHQ-4	0,486 ^b	0,001
PHQ-5	0,769 ^c	0,000
PHQ-6	0,176 ^a	0,384
PHQ-7	0,423 ^b	0,027
PHQ-8	0,280 ^d	0,116
PHQ-9	0,636 ^c	0,000

^a Acuerdo ligero.

^b Acuerdo moderado.

^c Acuerdo grande.

^d Acuerdo razonable.

$p = 0,109$). En el GHQ-12 se encontraron diferencias significativas en las 2 versiones ($Z = -2,294$; $p = 0,022$).

Consistencia interna

Los coeficientes alfa de Cronbach para las escalas GHQ-12, WHO-5 y PHQ-9 en lápiz y papel fueron de 0,768, 0,881 y 0,655, respectivamente, y para las GHQ-12, WHO-5 y PHQ-9 electrónicas fueron de 0,901, 0,872 y 0,836.

Fiabilidad test-retest

Los CCI entre las versiones en lápiz y papel y las electrónicas de las distintas escalas fueron de 0,616 ($p = 0,002$) para GHQ-12, 0,594 ($p = 0,001$) para WHO-5 y 0,584 ($p = 0,011$) para PHQ-9.

Equivalencia entre los formatos en lápiz y papel y electrónicos

Para contrastar los modelos en lápiz y papel y electrónicos calculamos el IAC y el correspondiente error medio cuadrático de aproximación (EMCA).

Para el GHQ-12 en lápiz y papel el IAC fue de 0,420 y el EMCA de 0,252, y para el GHQ-12 electrónico el IAC fue de 0,420 y el EMCA de 0,257. Para el WHO-5 en lápiz y papel el IAC fue de 0,888 y el EMCA de 0,216, y para el WHO-5 electrónico el IAC fue de 0,888 y el EMCA de 0,213. Para el PHQ-9 en lápiz y papel el IAC fue de 0,708 y el EMCA de 0,150, y para el PHQ-9 electrónico el IAC fue de 0,708 y el EMCA de 0,150.

Discusión

Salvo para el PHQ-9 en sus formatos en lápiz papel y electrónico-Likert, todos los instrumentos presentaron una consistencia interna suficiente para asegurar la fiabilidad de las escalas. Con respecto a la fiabilidad test-retest, aunque fue buena para los 2 formatos, resultó mayor en el paso al formato electrónico-Likert que en el paso al formato electrónico-slider. En el paso del formato lápiz y papel al formato electrónico-Likert, solo 2 ítems del GHQ-12 y uno del PHQ-9 tuvieron un acuerdo grande, siendo el acuerdo para

el resto de ítems bajo. La equivalencia entre los formatos en lápiz y papel y electrónico fue adecuada.

Nuestros resultados apoyan lo que han reflejado otros autores en trabajos previos que encuentran que si bien la equivalencia ente los cuestionarios en papel y sus versiones electrónicas es por lo general aceptable, esta migración no es completamente equivalente ni inmediata¹¹⁻¹³ y debe hacerse de acuerdo con unas normas consensuadas⁹.

Comparando la fiabilidad de nuestros resultados con las versiones previamente validadas en población española o de habla española —todas en versiones para completar en lápiz y papel—, podemos decir que la consistencia interna del cuestionario GHQ-12 es similar a la aportada por validaciones en población general española (alfa de Cronbach de 0,762 y 0,862); en nuestro caso todos los formatos tienen una buena consistencia interna, siendo especialmente alta la del formato electrónico-slider (alfa de Cronbach = 0,901). De modo similar, la consistencia interna de los distintos formatos del WHO-5 empleados en nuestro trabajo fue buena (con una alfa de Cronbach mayor de 0,80 en todos los casos) y estuvo al nivel de estudios previos, como un estudio europeo de población activa con parte de la muestra española (alfa de Cronbach = 0,82)²⁸ o un estudio en población española de más de 65 años (alfa de Cronbach = 0,86)²⁹. En el caso de la PHQ-9 no podemos comparar la consistencia interna con la de la validación española, pero es de interés señalar que es más baja que en los otros 2 cuestionarios; solo en el formato electrónico-slider el alfa de Cronbach es mayor de 0,80, mientras que en los demás ronda un valor de 0,60, claramente menor que la consistencia interna demostrada en poblaciones de habla española, con alfas de Cronbach de 0,89 en mujeres mexicanas³¹ o de 0,835 en pacientes chilenos de atención primaria³⁷.

Con respecto a las puntuaciones en los cuestionarios de nuestra muestra, tanto en el formato en lápiz y papel como en el electrónico, los participantes son catalogados como «sanos», aunque llama la atención que —a excepción del formato slider de la GHQ-12— tanto para la GHQ-12 como para la WHO-5 en los distintos formatos se acercan al punto de corte, para ser considerados casos, más de 2 de forma estricta en el caso de la GHQ (GHQ-0011) y menos de 13 para la WHO. Sin embargo, en la PHQ-9 no alcanzan el punto de corte de depresión en ninguno de los formatos. Aunque se aleja de nuestro tema de interés en este trabajo, estas puntuaciones ajustadas nos deberían hacer reflexionar sobre la autopercepción de salud y bienestar que tienen los estudiantes de ciencias de la salud; hay estudios que relacionan la sobrecarga académica con estrés, ansiedad y depresión³⁸, y los cuestionarios se administraron en el mes de mayo, tradicional mes de exámenes.

Pese al uso cada vez más extendido de formularios electrónicos, pocos son los trabajos que han estudiado su equivalencia como nos hemos planteado nosotros. En concreto, para los 3 cuestionarios que hemos usado, solo para el PHQ-9 un grupo ha estudiado previamente la adaptación electrónica; no hemos encontrado ningún trabajo similar para el WHO-5 ni para el GHQ-12, aunque sí para el GHQ-28. Con respecto al PHQ-9, en 2013, Bush et al.³⁹ compararon en 45 soldados en activo en una instalación militar de Estados Unidos las propiedades psicométricas de distintas medidas de salud —incluyendo el cuestionario PHQ-9— completados en papel, ordenador y smartphone. Como en nuestro caso,

las puntuaciones medias obtenidas en cada formato fueron similares y además en un rango parecido (en torno a una puntuación de 5). La consistencia interna que ellos encuentran es, sin embargo, mayor que la nuestra: 0,79 para el papel, 0,85 para el ordenador y 0,87 para el smartphone en su estudio frente a 0,58 para el papel (al compararlo con electrónico-Likert), 0,65 para el papel (al compararlo con electrónico-slider), 0,65 para el electrónico-Likert y 0,84 para el electrónico-slider. Es solo en este último formato en el que nuestras puntuaciones alcanzan las suyas, mientras que en el resto nuestra fiabilidad es baja. Del mismo modo, en este estudio encuentran una mejor fiabilidad test-retest que nosotros, con un CCI = 0,94 para papel-ordenador y 0,92 para papel-smartphone, claramente superior a los nuestros (0,68 para papel-Likert y 0,58 para papel-slider). Con respecto al GHQ-12, aunque no hemos encontrado ningún estudio que haya estudiado la validez de su formato electrónico, Vallejo et al.⁴⁰, en una muestra de 185 estudiantes de psicología en Madrid, estudiaron la validez del formato electrónico del GHQ-28, encontrando que ambos formatos son intercambiables. En su trabajo, reportan un alfa de Cronbach de 0,90 para los 2 formatos y una buena fiabilidad test-retest ($r = 0,69$). Nuestras alfas de Cronbach también son altas, siendo la del formato electrónico-slider (0,901) la que más se asemeja a las suyas. Nuestra fiabilidad test-retest es aún mejor que la suya, con un ICC = 0,802.

Como fortaleza, queremos destacar la novedad de nuestro trabajo, ya que aunque estamos asistiendo a la desaparición del papel son pocos los grupos que se han planteado estudiar la validez de los formatos electrónicos de los cuestionarios psicométricos más usados. Además, no hemos encontrado trabajos en los que se usen los cuestionarios que hemos empleado en formato visual analógico, el más apropiado en cuestionarios electrónicos^{15,16}. Pese a estas fortalezas, nuestro trabajo también tiene limitaciones. Algunas de ellas son el pequeño tamaño muestral y que se trata de una muestra no representativa, obtenida a conveniencia en una universidad española, por lo que no son resultados extrapolables a población general. Tampoco hemos tenido en cuenta el estatus socioeconómico, si los estudiantes tenían diagnósticos psiquiátricos u otros problemas de salud o eran consumidores de alguna sustancia; ni sabemos en qué soporte (web, teléfono, tableta, etc.) rellenaron las versiones electrónicas. Se hacen necesarios, por tanto, estudios futuros sobre este tema, con muestras de mayor tamaño y distintas procedencias.

Conclusión

Nuestros hallazgos demuestran que podemos asumir la equivalencia de los formatos electrónicos de los cuestionarios GHQ-12 y WHO-5, pero en el caso del PHQ-9 es necesaria prudencia. Además, los formatos slider se muestran como una alternativa válida a los cuestionarios tipo Likert en el entorno electrónico. El paso de un instrumento diseñado en papel a su uso en medio electrónico no es un proceso automático y no está exento de la necesidad de adaptación, de modo que cualquier instrumento que se cambie de soporte y método de aplicación debe ser validado previamente a su uso, más aún si se espera que tenga una aplicación clínica.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que en este artículo no aparecen datos de pacientes/participantes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes/participantes.

Financiación

El apoyo financiero para este trabajo fue proporcionado en parte por subvenciones de los proyectos ISCIII PI13/02200 FIS, del Plan Nacional de Drogas 2015I073 y de la beca Papiit IN108216. El acuerdo de financiación garantiza la independencia de los autores en el diseño del estudio, la interpretación de los datos, la escritura y la publicación del informe.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Escuela de Enfermería de la Fundación Jiménez Díaz. Universidad Autónoma de Madrid.

Anexo. Colaboradores del MEmind Study Group

Departamento de Psiquiatría, IIS-Fundación Jiménez Díaz, Madrid, España. Universidad Autónoma de Madrid, España: Irene Caro-Cañizares, Mónica Jiménez-Giménez, Juncal Sevilla-Vicente, Olga Bautista, Sara María Bañón-González, María Luz Palacios, María Natalia Silva, Jaime Chamorro-Delmo, Marta González-Granado, Sergio Sánchez-Alonso, Ernesto José Verdura-Vizcaíno, Miren Iza, Lucía Villoria-Borrego, Sonia Carollo-Vivian, Rocío Navarro-Jiménez, Laura Mata-Iturralde, Javier Fernández-Aurrecochea, Santiago Ovejero, Laura Muñoz-Lorenzo, Alba Rodríguez-Jover, Jorge Hernán Hoyos Marín, Carolina Vigil-López, Ana Rico-Romano, Rodrigo Carmona, Susana Amodeo-Escribano, Ana López-Gómez, Margarita Pérez-Fominaya, Covadonga Bonal-Giménez, Rosa Ana Bello-Sousa, Ruth Polo-del Río, Pedro Gutiérrez-Recacha, Iratxe Tapia-Jara, Marta Migoya-Borja, Elsa Arrua, Antonio Vian-Lains, Elena Hernando-Merino, Nora Palomar-Ciria, Leticia Serrano-Marugán, Alba Sedano-Capdevila, Marisa Herraiz, María Constanza Vera-Varela, Silvia Vallejo-Oñate.

Departamento de Psiquiatría, Hospital Universitario Infanta Elena, Valdemoro, Madrid, España: Rosana Codesal-Julián, Luis Sánchez-Pastor, Edurne Crespo-Llanos, Ainara Frade Ciudad, Marisa Martín-Calvo.

Departamento de Psiquiatría, Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, España: Laura de Andrés-Pastor, Pablo Puras-Rico, Miriam Agudo-Urbano, Diego

Laguna-Ortega, Sara Clariana-Martín, Eduardo Reguera-Nieto, Teresa Legido-Gil, María Guadalupe García-Jiménez, Raquel Álvarez-García, Pablo Portillo-de Antonio, Eva María Romero-Gómez Sara González-Granado.

Departamento de Psiquiatría, Hospital General de Villalba, Collado Villalba, Madrid, España: Ana Alcón-Durán, Juan Manuel García-Vega, Yago Cebolla-Meliá, Ezequiel Di Stasio, Pedro Martín-Calvo, Ana José Ortega.

Hospital 12 de Octubre, Madrid, España: Luis Agüera-Ortiz, Javier Rodríguez-Torresano, Javier Sanz-Fuentenebro, Miguel Ángel Jiménez-Arriero.

AGC Salud Mental, Área Sanitaria 3, Avilés, Asturias, España: Natalia Bretón-Díez, Juan José Martínez-Jambrina, Emilia García-Castro, María Fernández-Rodríguez, Mónica Álvarez-Villechenous.

Bibliografía

1. World Health Organization. The WHO-5 website [consultado 29 Ago 2015]. Disponible en: <https://www.psykiatri-regionh.dk/who-5/Pages/default.aspx>
2. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. *Med J Aust.* 2010;192 Suppl. 11:S53-6 [consultado 29 Ago 2015]. Disponible en: <https://www.mja.com.au/journal/2010/192/11/using-e-health-applications-deliver-newmental-health-services>
3. Mental Health Commission of Canada. E-mental health in Canada: Transforming the Mental Health System using technology. Ottawa, ON, 2014 [consultado 29 Ago 2015]. Disponible en: <http://www.mentalhealthcommission.ca>
4. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. *Qual Life Res.* 2015;24:1949-61.
5. Movsas B, Hunt D, Watkins-Bruner D, Lee WR, Tharpe H, Goldstein D, et al. Can electronic web-based technology improve quality of life data collection? Analysis of Radiation Therapy Oncology Group 0828. *Pract Radiat Oncol.* 2014;4:187-91.
6. Mulhern B, O’Gorman H, Rotherham N, Brazier J. Comparing the measurement equivalence of EQ-5D-5L across different modes of administration. *Health Qual Life Outcomes.* 2015;13:191.
7. Smith MJ, Reiter MJ, Crist BD, Schultz LG, Choma TJ. Improving patient satisfaction through computer-based questionnaires. *Orthopedics.* 2016;39:e31-5.
8. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value Health.* 2009;12:430-40.
9. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health.* 2009;12:419-29.
10. Rutherford C, Costa D, Mercieca-Bebber R, Rice H, Gabb L, King M. Mode of administration does not cause bias in patient-reported outcome results: A meta-analysis. *Qual Life Res.* 2016;25:559-74.
11. Alfnsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: A systematic review. *J Med Internet Res.* 2014;16:e268.
12. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported

- outcome measures: A meta-analytic review. *Value Health*. 2008;11:322–33.
13. 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.
 14. Hayes MH, Patterson DG. Experimental development of the graphic rating method. *Psychol Bull*. 1921;18:98–9.
 15. Funke F, Reips U-D. Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*. 2012, 1525822X12444061.
 16. Reips U-D, Funke F. Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behav Res Methods*. 2008;40:699–704.
 17. Rossi V, Pourtois G. Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS: A comparative review. *Anxiety Stress Coping*. 2012;25:603–45.
 18. Prensky M. Digital natives, digital immigrants. Part 1. *Horiz*. 2001;9:1–6.
 19. Catalá-López F, Hutton B, Page MJ, Vieta E, Tabarés-Seisdedos R, Moher D. Declaración de transparencia: un paso hacia la presentación completa de artículos de investigación. *Rev Psiquiatr Salud Ment (Barc.)*. 2016;9:63–4.
 20. Catalá-López F, Moher D, Tabarés-Seisdedos R. Improving transparency of scientific reporting to increase value and reduce waste in mental health research. *Rev Psiquiatr Salud Ment (Barc.)*. 2016;9:1–3.
 21. Goldberg, W. A user's guide to the General Health Questionnaire. NFER-Nelson; 1991 [consultado 16 Jun 2016]. Disponible en: <https://books.google.es/books?id=LpSuGQAACAAJ&dq=A+user%E2%80%99s+guide+to+the+General+Health+Questionnaire.&hl=es&sa=X&ved=0ahUKEwjLwq7nuazNAhXqBsAKHVFBCKQ6AEIHDA>
 22. Staehr JK. The use of well-being measures in primary health care — the DepCare project. En: *well-being measures in primary health care — the DepCare Project*. Geneva: World Health Organization, Regional Office for Europe; 1998.
 23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606–13.
 24. 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.
 25. 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.
 26. Sánchez-López MP, Dresch V. The 12-Item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema*. 2008;20:839–43.
 27. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: A systematic review of the literature. *Psychother Psychosom*. 2015;84:167–76.
 28. Boye K. Relatively different? How do gender differences in well-being depend on paid and unpaid work in Europe? *Soc Indic Res*. 2009;93:509–25.
 29. Lucas-Carrasco R. Reliability and validity of the Spanish version of the World Health Organization-Five Well-Being Index in elderly. *Psychiatry Clin Neurosci*. 2012;66:508–13.
 30. Encuesta europea sobre calidad de vida 2012. Eurofound [consultado 16 Jun 2016]. Disponible en: <http://www.eurofound.europa.eu/es/surveys/european-quality-of-life-surveys-eqls/european-quality-of-life-survey-2012>
 31. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review. *Gen Hosp Psychiatry*. 2010;32:345–59.
 32. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the Patient Health Questionnaire in diagnosing mental disorders in 1003 General Hospital Spanish inpatients. *Psychosom Med*. 2001;63:679–86.
 33. Familiar I, Ortiz-Panoso E, Hall B, Vieitez I, Romieu I, Lopez-Ridaura R, et al. Factor structure of the Spanish version of the Patient Health Questionnaire-9 in Mexican women. *Int J Methods Psychiatr Res*. 2015;24:74–82.
 34. IBM Downloading. IBM SPSS Statistics 23 — España; 2016 [consultado 10 Dec 2016]. Disponible en: <http://www.ibm.com/support, //www.ibm.com/support/docview.wss?uid=swg24038592>
 35. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105:399–412.
 36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159.
 37. Tomas Baader M, José Luis Molina F, Silvia Venezian B, Carmen Rojas C, Renata Farías S, Fierro-Freixenet C, et al. Validación y utilidad de la encuesta PHQ-9 (Patient Health Questionnaire) en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. *Rev Chil Neuro-Psiquiatr*. 2012;50:10–22.
 38. Paro HBMS, Morales NMO, Silva CHM, Rezende CHA, Pinto RMC, Morales RR, et al. Health-related quality of life of medical students. *Med Educ*. 2010;44:227–35.
 39. Bush NE, Skopp N, Smolenski D, Crumpton R, Fairall J. Behavioral screening measures delivered with a smartphone app: Psychometric properties and user preference. *J Nerv*. 2013;201:991–5.
 40. Vallejo MA, Jordán CM, Díaz MI, Comeche MI, Ortega J. Psychological assessment via the internet: A reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Checklist-90-Revised (SCL-90-R). *J Med Internet Res*. 2007;9:e2.