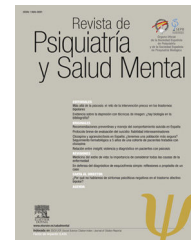




# Revista de Psiquiatría y Salud Mental

[www.elsevier.es/saludmental](http://www.elsevier.es/saludmental)



## ORIGINAL ARTICLE

# Comparative study of pencil-and-paper and electronic formats of GHQ-12, WHO-5 and PHQ-9 questionnaires<sup>☆</sup>



María Luisa Barrigón<sup>a</sup>, Ana María Rico-Romano<sup>a</sup>, Marta Ruiz-Gomez<sup>b</sup>,  
David Delgado-Gomez<sup>c</sup>, Igor Barahona<sup>d</sup>, Fuensanta Aroca<sup>d</sup>,  
Enrique Baca-García<sup>a,b,e,f,g,h,i,\*</sup>, MEMind Study Group<sup>abhijsk◇</sup>

<sup>a</sup> Departamento de Psiquiatría, Hospital Universitario Fundación Jiménez Díaz, IIS-Fundación Jiménez Díaz, Madrid, Spain

<sup>b</sup> Departamento de Psiquiatría, Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, Spain

<sup>c</sup> Departamento de Estadística, Universidad Carlos III, Getafe, Madrid, Spain

<sup>d</sup> Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico

<sup>e</sup> CIBERSAM, Madrid, Spain

<sup>f</sup> Universidad Autónoma de Madrid, Madrid, Spain

<sup>g</sup> Psychiatry Department, Columbia University, Nueva York, United States

<sup>h</sup> Departamento de Psiquiatría, Hospital Universitario Infanta Elena, Valdemoro, Madrid, Spain

<sup>i</sup> Departamento de Psiquiatría, Hospital General de Villalba, Collado Villalba, Madrid, Spain

<sup>j</sup> Hospital 12 de Octubre, Madrid, Spain

<sup>k</sup> AGC Salud Mental, Área Sanitaria 3, Avilés, Asturias, Spain

Received 26 July 2016; accepted 20 December 2016

Available online 8 June 2017

## KEYWORDS

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

## 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 24 h they completed their electronic versions via the web site [www.memind.net](http://www.memind.net). An electronic-Likert format was used by 24 participants, and 23 used an electronic-slider format. Internal consistency was measured by  $\alpha$ -Cronbach index and omega coefficient, and test-retest was measured by the intraclass correlation coefficient (ICC). Agreement between individual items was compared using weighted

<sup>☆</sup> Please cite this article as: Barrigón ML, Rico-Romano AM, Ruiz-Gomez M, Delgado-Gomez D, Barahona I, Aroca F, et al. Estudio comparativo de los formatos en lápiz y papel y electrónicos de los cuestionarios GHQ-12, WHO-5 y PHQ-9. Rev Psiquiatr Salud Ment (Barc). 2017;10:160–167.

\* Corresponding author.

E-mail address: [ebacgar2@yahoo.es](mailto:ebacgar2@yahoo.es) (E. Baca-García).

◇ MEMind Study Group collaborators are provided in Annex.

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 SEPB. Published by Elsevier España, S.L.U. All rights reserved.

## PALABRAS CLAVE

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

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

### 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 24 h en su versión electrónica a través de la web [www.memind.net](http://www.memind.net) (24 en formato electrónico-Likert y 23 en formato electrónico-slider). Se midió la consistencia interna mediante el índice  $\alpha$ -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 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 ( $\kappa$  ponderada = 0,143;  $p = 0,384$ ) hasta acuerdos altos en el ítem PHQ-5 ( $\kappa$  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 SEPB. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

## Introduction

In recent years telecommunications and computer technologies have entered the field of healthcare, giving rise to a new discipline that has been called "telemedicine".<sup>1</sup> Telemedicine has also reached mental health and is now a growth area.<sup>2</sup> A field of special interest is the evaluation of symptoms using electronic tools such as computers, smartphones or wearables,<sup>3</sup> so that what was classically studied in the psychometric field using "pencil-and-paper" now uses an electronic format. Different studies have found that participants feel more comfortable with electronic

questionnaires, answering the questions more easily and losing fewer data than occurs with paper questionnaires.<sup>4-7</sup>

When psychometric instruments are used in populations with characteristics different from the ones they were created for, a validation and adaptation process is necessary.<sup>8</sup> Likewise, in the migration from pencil-and-paper format to an electronic one, some initiatives attempt to show that both formats are equivalent instead of directly assuming that they are so.<sup>9</sup>

At the current time the literature supports the idea that pencil-and-paper questionnaires and electronic ones will supply equivalent data in patient-reported measurements

of the results in clinical trials<sup>4,10</sup> as well as when they are used as psychometric instruments. Thus three reviews study the comparative reliability of the pencil-and-paper and electronic versions of approximately 50 psychometric instruments — fundamentally for anxiety and depression — finding good reliability in general, although with discrepancies in some questionnaires.<sup>11–13</sup>

When psychometric questionnaires measure degree of agreement with a phrase, the answers are often shown in a Likert-type scale, with a certain number of options from greater to lesser degree of agreement. Analogue visual scales are an alternative to these questionnaires, as described by Hayes and Patterson in 1921.<sup>14</sup> In these the interviewee expresses his degree of agreement with the question on a lineal graph. The characteristics of the electronic format facilitate using an analogue visual format like a line, along which the interviewee can select the point that best represents his answer by sliding the mouse (the “slider”). Analogue visual or slider-type scales have been shown to be a good alternative in electronic/online questionnaires<sup>15,16</sup> and they make them simpler and faster to administer as well as easier to understand.<sup>17</sup>

A specific area of psychometrics consists of using self-administered questionnaires to screen for mental disorders. In the age of telemedicine having reliable electronic versions of instruments of this type would make it easier for a population increasingly familiar with technology to access them. This is especially important if we take the population of digital natives into account.<sup>18</sup> They are starting to use health services, while the majority of the professionals who deal with them are digital immigrants.<sup>18</sup>

Our aim here is to analyse whether the screening questionnaires in the General Health Questionnaire (GHQ-12), the Well-Being Index of the World Health Organisation (WHO-5) and the Patient Health Questionnaire (PHQ-9) have the same psychometric properties as the pencil-and-paper version in their electronic-Likert and electronic-slider forms. Our hypothesis is that the pencil-and-paper and electronic versions of each one of these scales will be interchangeable.

## Material and methods

### Participants and procedure

The sample consisted of 47 non-clinical volunteers aged from 18 to 24 years old. The majority of them were women (35 in all), and they were divided into 2 groups of 24 and 23 participants. All of them were Nursing Degree students in the Universidad Autónoma, Madrid.

In May 2016 the participants answered a self-administered interview that included the paper version of the following screening instruments: GHQ-12, WHO-5 and PHQ-9, in their corresponding classrooms after the last class in the morning. All of the students consented to take part. The questionnaires were completed anonymously, and no reward of any type was given. After the students had returned the questionnaires to the researcher, they were given access keys to the multi-platform tool [www.memind.net](http://www.memind.net), where they filled out the electronic versions of the same questionnaires in the following 24 h.

24 filled out an electronic-Likert version and 23 filled out an electronic-slider version.

This study was approved by the clinical research ethics committee of the Hospital Universitario Fundación Jiménez Díaz. After being informed of the nature of the study, all of the participants gave their informed consent in writing prior to taking part. The results have been presented following the transparency declaration proposed by Catalá-López et al.<sup>19,20</sup>

### Instruments

The questionnaires used were the 12-item version of the GHQ-12,<sup>21</sup> the WHO-5<sup>22</sup> and the PHQ-9 depression screening questionnaire.<sup>23</sup>

The participants first filled out the original versions in pencil-and-paper and then subsequently completed the electronic versions—developed by us—in electronic-Likert and electronic-slider formats.

#### The General Health Questionnaire (GHQ-12)

This is a self-administered questionnaire, designed for use in clinical circles to detect individuals with possible psychiatric disorders. It is composed of 12 items, 6 formulated positively and 6 negatively, which are answered on a Likert-type 4 point scale. The replies, which score from 0 to 3, are: “more than usual”, “the same as usual” and “much less than usual” for the positive questions, and “no, not at all”, “no more than usual” “somewhat more than usual” and “much more than usual” for the negative items. Higher scores indicate worse health. Different models have been developed for scoring: the standard method (GHQ-0011), corrected scoring (GHQ-0111) and Likert scoring (GHQ-0123).

The standard method is considered to be the most suitable for case identification; this gives a score of from 0 to 12, with the cut-off point located between 2 and 3<sup>24</sup>; this was the method used in the comparison between the pencil-and-paper and electronic-Likert formats. The GHQ-0123 score was used for the group of participants which compared the pencil-and-paper version and the electronic-slider one.

The GHQ-12 has been validated in 15 countries and it achieves good reliability, with Cronbach’s alphas from 0.82 to 0.86.<sup>25</sup> In the Spanish population, different studies have found good reliability in the general population, with Cronbach’s alphas from 0.76<sup>26</sup> to 0.86.<sup>25</sup>

#### The WHO well-being index (WHO-5)

This self-administered scale with 5 items is used to measure the feeling of well-being. The 5 phrases it contains have the particularity that they are all expressed positively. They are: (1) “I feel happy and in a good mood”, (2) “I feel peaceful and relaxed”, (3) “I have felt active and energetic”, (4) “I woke up fresh and rested”, and (5) “My everyday life has been full of things which interest me”. It is available gratis in several languages at <http://www.who-5.org/>. For each one of the items, the degree of well-being is scored for the 2 previous weeks on a Likert-type scale from 0 (never) to 5 (all of the time); the total score varies from 0 to 25, so that the highest scores are associated with a stronger feeling of well-being, while scores below 13 have been linked to depression.<sup>27</sup>

The WHO-5 is a scale used around the world and not only in the field of mental health, but also in general healthcare. It has been proven to be suitably valid as a screening tool for depression and as a means of measuring evolution in several clinical trials.<sup>27</sup> Specifically, in a study of an active adult European population which included Spanish population, its internal consistency amounted to 0.82 (Cronbach's alpha),<sup>28</sup> while in a Spanish population over the age of 65 years old its internal consistency was 0.86.<sup>29</sup> In the last available European quality of life survey, in 2012, the average score in Spain was 65.4.<sup>30</sup>

### The Patient-9 Health Questionnaire (PHQ-9)

This is a self-administered questionnaire that is used for depression screening. It is composed of 9 items formulated based on the diagnostic criteria for depression of the DSM-IV; each item scores from 0 to 3 (0 = no day; 1 = several days; 2 = more than half of the days, and 3 = almost every day). Although there are other ways of scoring, when it is used as a depression screening tool the total score of the PHQ-9 is calculated by adding up the score for each item, and it varies from 0 to 27; scores above 10 indicate moderate to high levels of depression.<sup>23</sup>

The usefulness of the PHQ-9 as a depression screening tool has been shown to be equal to or even greater than those of other measures of depression, and it obtained a good level of validity in different studies.<sup>31</sup> More specifically, Kroenke et al.<sup>23</sup> reported an internal consistency for the PHQ-9 of from 0.86 to 0.89. In Spain, Díez-Quevedo et al. validated the PHQ-9 in 1003 hospitalised patients, showing good agreement between the diagnoses using the PHQ-9 and clinical diagnoses ( $\kappa = 0.74$ ; total precision, 88%; sensitivity, 87%; specificity, 88%).<sup>32</sup> in the Spanish-speaking population the validation study in the Mexican population stands out; in 55,000 women in the Mexican Teachers' Cohort the questionnaire was found to have a high level of internal consistency, with a Cronbach's alpha of 0.89.<sup>33</sup>

### MEmind Wellness Tracker

The MEmind Wellness Tracker tool was used for the electronic administration of the scales. It was developed in the Psychiatry Department of the Hospital Universitario Fundación Jiménez Díaz. This web application is available at [www.memind.net](http://www.memind.net), and it works with all types of internet-access devices (computers, tablets and smartphones) using any operating system. The website has 2 interfaces, one for the researcher and the other for the user. Once they had given the researcher the printed questionnaire, the students were registered in the website and given access keys (users and password), after which the researcher randomly assigned them one of the 2 types of electronic questionnaire programmed to be completed in the following 24 h.

After accessing the platform and filling out the questionnaires, the data were stored in a secure server and encrypted using Secure Socket Layer/Transport Layer Security (SSL/TLS). Only the head researcher (EBG) has access to the server. MEmind uses encrypted keys with 256 bit codes based on the AES-256 algorithm. An external auditor guarantees that these levels of security comply with the maximum requisites of the data protection law.

## Statistical analysis

Statistical analysis was performed using version 23.0 of the SPSS package.<sup>34</sup> Firstly, given the small size of the sample and the absence of normality, the scores were compared in each one of the scales in the pencil-and-paper and electronic versions using Wilcoxon's signed range test. Then the reliability was calculated using consistency and agreement indexes. Cronbach's alpha was used to measure the internal consistency of the questionnaires in both formats. Additionally, and given the limitations of the alpha coefficient pointed out by several authors, the omega coefficient was also calculated.<sup>35</sup> The intraclass correlation coefficient (ICC) was used to estimate test-retest reliability; ICC values under 0.4 indicate low reliability, at from 0.4 to 0.75 they indicate from mediocre to good reliability, and above 0.75 they indicate excellent reliability.<sup>33</sup> The degree of agreement between the different items in the pencil-and-paper and electronic-Likert versions of the questionnaires was measured using weighted kappa coefficients, and to interpret them the criteria proposed by Landis and Koch were used, according to which values lower than 0 indicate a poor level of agreement, values from 0.01 to 0.20 indicate slight agreement, from 0.21 to 0.40 a reasonable level of agreement, a moderate agreement from 0.41 to 0.60, from 0.61 to 0.80 a strong agreement and from 0.81 to 1.00 an almost perfect agreement.<sup>36</sup> Lastly, to compare the pencil-and-paper and electronic models, and thereby compare their dimensional structure, we calculate the comparative fit index (CFI) for the 3 scales evaluated.

## Results

### The switch from a pencil-and-paper format to an electronic-Likert one

There were no differences between the pencil-and-paper and electronic versions in any of the scales: GHQ-12 ( $Z = -1.709$ ;  $P = .087$ ); WHO-5 ( $Z = -1.067$ ;  $P = .286$ ) and PHQ-9 ( $Z = -0.199$ ;  $P = .842$ ), with the following average scores for each one of them:  $1.96 \pm 2.66$  in pencil-and-paper vs  $1.50 \pm 2.55$  in electronic format for the GHQ-12;  $13.63 \pm 2.66$  in pencil-and-paper vs  $14.42 \pm 4.03$  in electronic format for the WHO-5, and  $4.87 \pm 2.80$  in pencil-and-paper vs  $4.83 \pm 2.85$  in electronic format for the PHQ-9.

### Internal consistency

Cronbach's alpha coefficients for the GHQ-12, WHO-5 and PHQ-9 in pencil-and-paper were 0.872, 0.774 and 0.586, respectively, and for the electronic GHQ-12, WHO-5 and PHQ-9 they were 0.835, 0.835 and 0.654. Additionally, the omega coefficients for the GHQ-12, WHO-5 and PHQ-9 in pencil-and-paper were 0.900, 0.900 and 0.720, respectively, and for the electronic versions of the GHQ-12, WHO-5 and PHQ-9 they were 0.890, 0.900 and 0.730.

### Test-retest reliability

The ICC between the pencil-and-paper and electronic versions of the different scales were 0.802 ( $P < .001$ ) for the

**Table 1** Agreement between the items for the GHQ-12 questionnaire.

Item	Weighted Kappa	<i>P</i>
GHQ-1	0.478 <sup>a</sup>	0.001
GHQ-2	0.556 <sup>a</sup>	0.000
GHQ-3	0.492 <sup>a</sup>	0.000
GHQ-4	0.273 <sup>b</sup>	0.094
GHQ-5	0.400 <sup>b</sup>	0.001
GHQ-6	0.670 <sup>c</sup>	0.000
GHQ-7	0.509 <sup>a</sup>	0.000
GHQ-8	0.635 <sup>c</sup>	0.000
GHQ-9	0.563 <sup>a</sup>	0.000
GHQ-10	0.613 <sup>c</sup>	0.001
GHQ-11	0.774 <sup>c</sup>	0.000
GHQ-12	0.373 <sup>b</sup>	0.011

<sup>a</sup> Moderate agreement.

<sup>b</sup> Reasonable agreement.

<sup>c</sup> Strong agreement.

GHQ-12, 0.726 ( $P < .001$ ) for the WHO-5 and 0.682 ( $P < .001$ ) for the PHQ-9.

#### Agreement between individual items

In the GHQ-12 questionnaire the agreement between the items was reasonable at the least and even high for several of them (Table 1). For the WHO-5, except for item 3 that was reasonable, the agreement between the items was moderate (Table 2). Finally, for the PHQ-9, except for items PHQ-1 and PHQ-6, in which the level of agreement was low, the level of agreement was at least reasonable and even high for items PHQ-5 and PHQ-9 (Table 3).

#### The switch from the pencil-and-paper format to electronic-slider

There were no statistically significant differences between both versions when the WHO-5 and PHQ-9 scales were compared: WHO-5 ( $Z = -0.974$ ;  $P = .330$ ) and PHQ-9 ( $Z = -1.601$ ;  $P = .109$ ). Statistically significant differences were found between both versions of the GHQ-12 ( $Z = -2.294$ ;  $P = .022$ ).

#### Internal consistency

The Cronbach's alpha coefficients for the GHQ-12, WHO-5 and PHQ-9 scales in pencil-and-paper were 0.768, 0.881 and 0.655, respectively, and for the electronic versions of the

**Table 2** Agreement between the items for the WHO-5 questionnaire.

Item	Weighted Kappa	<i>P</i>
WHO-1	0.565 <sup>a</sup>	0.000
WHO-2	0.451 <sup>a</sup>	0.001
WHO-3	0.358 <sup>b</sup>	0.004
WHO-4	0.481 <sup>a</sup>	0.000
WHO-5	0.571 <sup>a</sup>	0.000

<sup>a</sup> Moderate agreement.

<sup>b</sup> Reasonable agreement.

**Table 3** Agreement between the items for the PHQ-9 questionnaire.

Item	Weighted Kappa	<i>P</i>
PHQ-1	0.143 <sup>a</sup>	0.384
PHQ-2	0.448 <sup>b</sup>	0.006
PHQ-3	0.516 <sup>b</sup>	0.000
PHQ-4	0.486 <sup>b</sup>	0.001
PHQ-5	0.769 <sup>c</sup>	0.000
PHQ-6	0.176 <sup>a</sup>	0.384
PHQ-7	0.423 <sup>b</sup>	0.027
PHQ-8	0.280 <sup>d</sup>	0.116
PHQ-9	0.636 <sup>c</sup>	0.000

<sup>a</sup> Slight agreement.

<sup>b</sup> Moderate agreement.

<sup>c</sup> Strong agreement.

<sup>d</sup> Reasonable agreement.

GHQ-12, WHO-5 and PHQ-9 scales they were 0.901, 0.872 and 0.836.

#### Test-retest reliability

The ICC between the pencil-and-paper and electronic versions of the different scales was 0.616 ( $P = .002$ ) for the GHQ-12, 0.594 ( $P = .001$ ) for the WHO-5 and 0.584 ( $P = .011$ ) for the PHQ-9.

#### Equivalence between the pencil-and-paper and electronic formats

To compare the pencil-and-paper and electronic models we calculate the CFI and the corresponding mean squared error of approximation (MSEA).

For the pencil-and-paper GHQ-12 the CFI was 0.420 and the MSEA was 0.252, and for the electronic GHQ-12 the CFI was 0.420 and the MSEA was 0.257. For the pencil-and-paper WHO-5 the CFI was 0.888 and the MSEA was 0.216, and for the electronic WHO-5 the CFI was 0.888 and the MSEA was 0.213. For the pencil-and-paper PHQ-9 the CFI was 0.708 and the MSEA was 0.150, and for the electronic PHQ-9 the CFI was 0.708 and the MSEA was 0.150.

#### Discussion

Except for the pencil-and-paper and electronic-Likert PHQ-9, all of the instruments presented sufficient internal consistency to ensure the reliability of the scales. Respecting test-retest reliability, although this was good for both formats it was greater in the electronic-Likert format than it was in the switch to the electronic-slider format. In the switch from the pencil-and-paper format to electronic-Likert, only 2 items in the GHQ-12 and one in the PHQ-9 achieved a high level of agreement while the level of agreement was low for the other items. The equivalence between the pencil-and-paper and electronic formats was sufficient.

Our results support the views of other authors in previous works, who found that although the level of equivalence between pencil-and-paper and electronic questionnaires is acceptable in general, this migration is neither completely



equivalent nor immediate,<sup>11–13</sup> and it must take place using certain agreed norms.<sup>9</sup>

Comparing the reliability of our results with the previous validated versions in Spanish populations or Spanish speaking ones — all of the versions to be completed using pencil-and-paper — we can state that the internal consistency of the GHQ-12 questionnaire is similar to that of the validations in the general Spanish population (Cronbach's alpha at 0.762 and 0.862); in our case all of the formats have good internal consistency, while the electronic-slider one is especially good in this respect (Cronbach's alpha = 0.901). Similarly, the internal consistency of the different formats of the WHO-5 questionnaire used in our study was good (Cronbach's alpha higher than 0.80 in all cases) and it was at the same level as the previous studies, with a European study of active population with part of the Spanish sample (Cronbach's alpha = 0.82)<sup>28</sup> or a study in a Spanish population over the age of 65 years old (Cronbach's alpha = 0.86).<sup>29</sup> In the case of the PHQ-9 we are unable to compare the internal consistency with that corresponding to the Spanish validation, but it is of interest to point out that it is lower than in both of the other questionnaires; Cronbach's alpha is only higher than 0.80 for the electronic-slider format, while for the others it stands at a value of around 0.60, which is clearly lower than the internal consistency found in Spanish-speaking populations, with a Cronbach's of 0.89 in Mexican women<sup>31</sup> or 0.835 in primary care patients in Chile.<sup>37</sup>

Respecting the scores in the questionnaires of our sample in pencil-and-paper as well as electronic formats, the participants are classified as "healthy". Nevertheless, it is striking that (except for the slider format GHQ-12—the different formats of the GHQ-12 and WHO-5 approach the cut-off point to be considered cases: strictly more than 2 for the GHQ (GHQ-0011) and less than 13 for the WHO. However, in the PHQ-9 they do not reach the cut-off point for depression in any of the formats. Although it is outside our field of interest for this study, these indicative scores should lead us to reflect on the self-perception of health and well-being felt by health sciences students: some studies link academic overload with stress, anxiety and depression,<sup>38</sup> and these questionnaires were administered in the month of May, which is traditionally exam time.

In spite of the increasingly widespread use of electronic forms, few studies have researched their equivalence as we do here. More specifically, of the 3 questionnaires that we used, only for the PHQ-9 has one group studied its electronic version; we found no similar studies for the WHO-5 or the GHQ-12, although we did for the GHQ-28. Respecting the PHQ-9, in 2013 Bush et al.<sup>39</sup> compared the psychometric properties of different health measurements in 45 serving soldiers in a United States military installation. These included the PHQ-9 questionnaire, and they were completed in paper, computer and smartphone. As in our case, the average scores obtained in each format were similar, and they were also in a similar range (around a score of 5). Nevertheless, the internal consistency that they found is greater than our figures: 0.79 for paper, 0.85 for computers and 0.87 for smartphones in their study. This compares with our figures of 0.58 for paper (compared with electronic-Likert), 0.65 for paper (compared with the electronic-slider), 0.65 for the

electronic-Likert and 0.84 for the electronic-slider. It is only for the latter format that our scores reach the same level as theirs, while our reliability is low for the others. Thus in this study they found a higher level of test-retest reliability than we did, with an ICC = 0.94 for paper-computer and 0.92 for paper-smartphone, which are clearly higher than our figures (0.68 for paper-Likert and 0.58 for paper-slider). Respecting the GHQ-12, although we found no study of the validity of its electronic format, in a sample of 185 psychology students in Madrid Vallejo et al.<sup>40</sup> studied the validity of the electronic format of the GHQ-28, and they found both formats are interchangeable. They report a Cronbach's alpha of 0.90 for both formats and good test-retest reliability ( $r = 0.69$ ). Our Cronbach's alphas are high too, and the electronic-slider format (0.901) is the one that is closest to theirs. Our test-retest reliability is even higher than theirs, with an ICC = 0.802.

We would like to underline the novelty of our work as one of its strong points, as although paper is disappearing few groups have considered studying the validity of the electronic versions of the most widely used psychometric questionnaires. Moreover, we found no works that studied the questionnaires we used in analogue visual format, which is the most suitable for electronic questionnaires.<sup>15,16</sup> In spite of these strong points, our work also has limitations. Some of these are its small sample size and the non-representative nature of the same, as it was conveniently recruited in a Spanish university. These results can therefore not be extrapolated to the general population. Nor did we take socio-economic status into account, or whether the students had psychiatric diagnoses or other health problems, or whether they consumed substances; nor do we know the support that they used (web, telephone or tablet, etc.) to fill out the electronic versions. Further studies are therefore required in the future with larger and more varied samples.

## Conclusions

Our findings show that we are able to assume the equivalence of the electronic formats of the GHQ-12 and WHO-5 questionnaires, although in the case of the PHQ-9 prudence is required. Moreover, slider formats were found to be a valid alternative to Likert-style questionnaires in the electronic environment. The switch from an instrument designed on paper to using it in an electronic medium is not an automatic process that requires no adaptation, so that every instrument which migrates to another support and method of administration must be validated before it is used, more so if it is expected to be used clinically.

## Ethical responsibilities

**Protection of people and animals.** The authors declare that no experiments were conducted in human beings or animals for this research.

**Confidentiality of data.** The authors declare that no patient or participant data appear in this paper.

**Right to privacy and informed consent.** The authors declare that no patient or participant data appear in this paper.

## Financing

The financial support for this work was supplied in part by subsidies from the ISCIII PI13/02200 FIS projects, the National Drugs Plan 20151073 and the Papiit IN108216 grant. The financing agreement guarantees the independence of the authors in the study design, the interpretation of data and the writing and publication of the report.

## Conflict of interests

The authors have no conflict of interests to declare.

## Acknowledgements

The Nursing School of the Fundación Jiménez Díaz. Universidad Autónoma de Madrid.

## Annex. MEmind Study Group collaborators

*Psychiatry Department, IIS-Fundación Jiménez Díaz, Madrid, Spain. Universidad Autónoma de Madrid, Spain:* 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-Aurrecoechea, 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 Rio, 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.

*Psychiatry Department, Hospital Universitario Infanta Elena, Valdemoro, Madrid, Spain:* Rosana Codesal-Julián, Luis Sánchez-Pastor, Edurne Crespo-Llanos, Ainara Frade Ciudad, Marisa Martín-Calvo.

*Psychiatry Department, Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, Spain:* 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.

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

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

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

## References

1. World Health Organization. The WHO-5 website. Available from: <https://www.psykiatri-regionh.dk/who-5/Pages/default.aspx> [accessed 29.08.15].
2. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. *Med J Aust.* 2010;192 Suppl. 11:S53–6. Available from: <https://www.mja.com.au/journal/2010/192/11/using-e-health-applications-deliver-newmental-health-services> [accessed 29.08.15].
3. Mental Health Commission of Canada. E-mental health in Canada: transforming the mental health system using technology; 2014. Ottawa, ON. Available from: <http://www.mentalhealthcommission.ca> [accessed 29.08.15].
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. *On the Horizon*. 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. Available from: <https://books.google.es/books?id=LpSuGQAACAAJ&dq=A+user%E2%80%99s+guide+to+the+General+Health+Questionnaire.&hl=es&sa=X&ved=0ahUKEwjLwq7nuazNAhXqBsAKHVFCMkQ6AEIHDA&pg=PA1> [accessed 16.06.16].
22. Staehr JK. The use of 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. Available from: <http://www.eurofound.europa.eu/es/surveys/european-quality-of-life-surveys-eqls/european-quality-of-life-survey-2012> [accessed 16.06.16].
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. Díez-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-Panozo 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. Available from: <http://www.ibm.com/support,http://www.ibm.com/support/docview.wss?uid=swg24038592> [accessed 10.12.16].
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 Ment Dis*. 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 Check-List-90-Revised (SCL-90-R). *J Med Internet Res*. 2007;9:e2.