

GHQ-12 in adolescents: contributions to the controversial factorial validity

Eduardo J. Pedrero-Pérez¹, Carlos Mora-Rodríguez², Rosana Rodríguez-Gómez¹, María Teresa Benítez-Robredo¹, Ana Ordoñez-Franco², Lidio González-Robledo³, and Susana Méndez-Gago³

¹ Research and Training Unit. Assessment and Quality Department. Madrid Salud. Madrid City Council. (Spain).

² Prevention Department. Institute of Addictions. Madrid Salud. Madrid City Council (Spain).

³ Faculty of Psychology. Camilo José Cela University, Madrid (Spain).

Título: GHQ-12 en adolescentes: contribuciones a la controversia sobre su validez factorial.

Resumen: *Objetivo:* El General Health Questionnaire-12 (GHQ-12) es de los cuestionarios más utilizados en la población adulta, pero hay pocos estudios que exploren su aplicabilidad en adolescentes. El principal problema es la diversidad de resultados en su estructura factorial, ya que diferentes estudios han encontrado una, dos o tres dimensiones. El propósito de este estudio fue explorar la estructura factorial del GHQ-12 en adolescentes. *Método:* Estudio descriptivo observacional multicéntrico transversal para centros educativos de la ciudad de Madrid. Se estratificó una muestra aleatoria de $n = 2171$ jóvenes de 14-16 años según el nivel de desarrollo y financiación de la escuela (pública, privada o mixta). Se aplicó un análisis factorial sin restricciones basado en la matriz de correlaciones policóricas o tetracóricas (dependiendo del método de puntuación de la prueba). *Resultados:* El análisis paralelo optimizado proporcionó una estructura inequívocamente unifactorial. El 30% de los varones y el 42% de las adolescentes presentaban riesgo de mala salud mental. *Conclusiones:* El GHQ-12 es un test unidimensional para el cribado del riesgo de mala salud mental, con excelentes propiedades psicométricas para su aplicación en población adolescente. El uso de métodos estadísticos adecuados puede superar viejas controversias y promover la correcta interpretación de los resultados de la prueba.

Palabras clave: GHQ-12; salud mental; adolescencia; cribado; análisis factorial; detección precoz; cuestionario.

Abstract: *Objective:* The General Health Questionnaire-12 (GHQ-12) is one of the most widely used questionnaires in the adult population, but there are not as many studies exploring its applicability in adolescents. The main problem is the diversity of results on its factorial structure, as different studies have shown one, two or three dimensions. The purpose of this study was to explore the GHQ-12 factorial structure for adolescents. *Method:* An observational multicenter cross-sectional descriptive study was designed for schools in the city of Madrid. A random sample of $n=2,171$, 14-16-year-olds was stratified by the school's development level and funding (public, private or mixed). Unrestricted factor analysis was applied based on the polychoric or tetrachoric correlations matrix (depending on the test scoring method). *Results:* The optimized parallel analysis provided an unequivocal one-factor structure, confirmed by unidimensionality assessment. Around 30% of male and 42% of female adolescents were found to be at risk of developing mental health problems. *Conclusions:* The GHQ-12 is a one-dimensional test for screening psychological distress, with excellent psychometric properties for its application in an adolescent population. The use of adequate statistical methods can overcome old controversies and promote proper application and interpretation of the results provided by the test.

Keywords: GHQ-12; mental health; adolescence; screening; factorial analysis; early detection; questionnaire.

Introduction

Early detection of psychological problems is, doubtless, a goal of enormous interest for diagnosis and early intervention. But it is still more important in adolescent populations, as early intervention can slow down the stabilization and progress of more severe psychopathology in later stages of life. This need has encouraged the use of screening instruments easily administered to large samples and with sufficient sensitivity and specificity to make early detection and specialized intervention possible (Sheldrick et al., 2015).

One of the most widely used tests in the world is the 12-item version of the General Health Questionnaire (GHQ-12), originally proposed by Goldberg (1972). The 12-item questionnaire asks about such matters as the ability to concentrate, sleeping problems, decision-making problems, perceived stress, self-confidence or perceived happiness. It does not enable the underlying psychopathology to be established but explores perceptions of stress that could be manifestations of an underlying disorder. There are studies backing

both its internal and external validity for detecting nonpsychotic psychopathology (Sánchez-López, & Dresch, 2008).

Although there is wide agreement about the ability of the test to detect such signs of distress, there is persistent controversy concerning what the test really measures. The main problem is that there is no agreement on the number of dimensions that configure the test, which is an indispensable requirement for interpreting the results of its application and comparing the results found. Among the studies published are those which propose that the GHQ-12 has a one, two or three-dimensional structure.

Confusion about the structure of GHQ-12 inevitably leads to misuse and the existence of works that interpret the results in very different ways. Thus, some studies explore the "prevalence" of specific mental health problems in adolescents, such as anxiety or depression, when GHQ-12 is merely a screening test that is unlikely to be able to measure specific disorders and in any way estimate prevalence (Bansal, Goyal, & Srivastava, 2009; Mann et al., 2011) or find links between such supposed disorders and other mental health problems (Almeida et al., 2019; Ogawa et al., 2019; Ojio, Nishida, Shimodera, Togo, & Sasaki, 2016). Therefore, it is necessary to know the true structure of the questionnaire, the dimensions it is able to measure and, ultimately, what it really measures.

When the test has been applied in adult populations, sev-

*** Correspondence address [Dirección para correspondencia]:**

Eduardo J. Pedrero-Pérez. Departamento de Evaluación y Calidad. Unidad Técnica de Formación e Investigación. Avenida del Mediterráneo 62, 28007, Madrid (Spain). E-mail: pedreropj@madrid.es
(Article received: 21-4-2019; revised: 23-7-2019; accepted: 20-11-2019)

eral possible structural formulas have been found (Shevlin, & Adamson, 2005), and in general, acceptable psychometric properties and diagnostic validity (Makowska et al., 2002). When the test has been used in samples of adolescents, two and three-factor solutions have predominated, as summa-

rized in Table 1. Through the statistical methods used, differences have been found in very diverse ethnic groups (Bowe, 2017). It may be observed that there are two studies in Spanish adolescents, one of which found a two-factor solution, and the other, a three-factor structure.

Table 1. Factorial studies of GHQ-12 in adolescents.

Author	Sample	Country	Method	Factorial solution
French, & Tait, 2004.	$n = 336$; Age 11-15.	Australia	CFA, weighted least squares	3
López-Castedo, & Fernández, 2005.	$n = 1,930$; Age 14-19.	Spain	PCA (Kaiser criterion, Scree-plot).	2
Sarkova, et al, 2006.	$n = 519$; Age 10-14.	Hungary and Slovakia	PCA (extraction criteria not specified)	2 y 3
Li, Chung, Chui, & Chan, 2009.	$n = 1,945$; Age 12-19.	China	PCA (Kaiser criterion, Scree-plot). CFA, weighted least squares	2
Padrón, et al., 2012.	$n = 4,146$; Age 13-18.	Spain	PCA (Kaiser criterion, Scree-plot). CFA, generalized weighted least squares	3
Glozah, & Pevalin, 2015.	$n = 770$; Mean age 16.9.	Ghana	EFA, Maximum Likelihood (Kaiser criterion, Scree-plot).	2

Note: CFA= Confirmatory Factor Analysis; PCA= Principal Component Analysis; EFA= Exploratory Factor Analysis.

One main reason justifying these different solutions has been suggested. There is empirical evidence that people do not answer positive responses (in the direction of distress) the same way as negative, generating two artefactual factors, the first of which is the more consistent of the two, because negative items are ambiguous about the nonexistence of such symptoms (Hankins, 2008a). The methods generally used for structural analysis are inadequate for the nature of the data explored. The use of the Kaiser criterion (Eigenvalue > 1) or Scree plots are inadequate to the extent that they assume that the Likert scales represent continuum scores and tend to overestimate the presence of artefactual dimensions (Ruscio, & Roche, 2012). The same is true of methods such as extraction of principal components, usually with the Pearson's correlation matrix. These methods contribute to grouping responses based on the distribution of the variables instead of by their content. Thus, the responses to questions leading to greater ambiguity may simply be grouped by the type of response, generating some spurious factors. The Cronbach's alpha coefficient is not an adequate estimator for studying the internal consistency of ordinal scales either, nor are Likert responses (Crutzen, & Peters, 2017; Yang, & Green, 2011). Several studies have shown that when these biases are controlled for and appropriate statistics are applied, the solution found is always one-factor (Hankins, 2008b; Rey et al., 2014; Romppel et al., 2013).

Another controversial question is whether the interpretation of the results of the GHQ-12 in adolescents should respond to the same rules as for adults, that is, place the cutoff point at three or more affirmative answers. There are no studies exploring the predictive validity of the GHQ-12 in Spanish adolescent populations. When it has been applied to samples from other countries, the data are very contradictory. Some suggest the scores of adolescents should be interpreted the same way as adults (French, & Tait, 2004), others use Likert scores, finding high prevalence of psychopathological problems (Bansal, Goyal, & Srivastava, 2009), or pro-

pose cutoff points of 10 for males and 11 for females on continuum scores (Baksheev et al., 2011). The author of the GHQ-12 suggested using the sample mean as the cutoff point when no previous data are available (Goldberg, & Williams, 1998).

The purpose of this study was to explore the dimensionality of the GHQ-12 in a representative sample of adolescents enrolled in schools in a large Western city (Madrid, Spain), using the most appropriate methods, able to control for the biases mentioned. In addition, the differences between males and females in the sample were estimated according to the factor solution found.

Method

Participants and procedure

Study population

The population enrolled in 4th year Obligatory Secondary Education in public, semi-private and private high schools in the city of Madrid during academic year 2016-2017.

Design

An observational multicenter cross-sectional descriptive study was designed for public, semi-private and private schools in the city of Madrid. The main objective of this study was to understand the relationships between the use and abuse of information and communication technologies and various health indicators, including the risks of poor mental health (Pedrero-Pérez et al., 2019). Stratified random sampling was applied by (a) the city districts grouped in four strata as used in the Study of Health in the City of Madrid 2014 (Díaz-Olalla, & Benítez-Robredo, 2015), based on municipal administrative division into the 21 districts which make up the municipality of Madrid, reflecting grouping

closely related to the level of development of the residential area; (b) school funding, based on the complete list of schools in the City of Madrid, including their funding (public, subsidized or private). Their location and the number of 4th year Obligatory Secondary Education students (ESO, usually 15-16-year-olds) per center, as provided by the education authority, the Council for Education and Research in the Community (Region) of Madrid.

The directors and guidance counselors of the thirty-four participating high schools were previously informed of study details and gave their consent. Fieldwork was done by professionals with previous experience in educational intervention who received training in the digital application of the questionnaire. The GHQ-12 questionnaire (among others) was uploaded to an online digital application (Google Form[®]), enabling simultaneous anonymous answers. The participants were also requested to enter their sex, age and school. The questionnaires were filled out in a computer lab in which each participant had a computer connected to the Internet. The instructors, both the school's teachers and visiting professionals, remained presented while it was being filled out. The Data were acquired from December 2016 to March 2017. Informed consent was previously requested from parents, tutors or legal guardians of the participants. Only students who handed in this signed consent participated in the study (9% excluded). The data collection method guaranteed anonymity of the participants. The study was approved by the General Direction of Nursery, Primary and Secondary Education of the Community of Madrid. Data collection was done during the month of December 2016 and the months of January, February and March 2017: 12 schools in December 2016, nine schools in January 2017, 12 schools in February 2017 and one in March 2017.

Participants

A representative sample was taken of the total population of 4th year ESO students ($n = 2,341$) at the 34 schools selected, stratified by level of development of the neighborhoods and ownership: public, semi-private or private. For this study, and to keep data homogeneous, only questionnaires answered by 15 or 16-year-old students were included ($n = 2,171$; 50.2% males). According to population distribution, 34.3% went to a public school, 58.1% to a semi-private one and 7.6% to a private one.

Instrument

The Spanish (Lobo, & Muñoz, 1996) 12-item version by Goldberg (1972) was used. It is a 12-item self-report questionnaire with multiple-choice answers, always with four choices. Six of the items are positive (psychological distress) and six are negative (no distress). When they are corrected, all the items are interpreted in the direction of distress, so the higher the scores, the greater the distress is. Two ways of correcting the responses have been suggested: (a) GHQ-Likert scores from 0 to 3 where higher scores are indicators of worse mental health, and (b) GHQ criterion score, assign-

ing values of 0-0-1-1 to the item choices. Criterion scores in over two affirmative choices suggest risk of poor mental health (Makowska et al., 2002).

Data analysis

First, outliers (participants who answered randomly or incongruently) were excluded after measuring the Mahalanobis distance ($p < .001$). The descriptive statistics were found for the items. The normality hypothesis was checked applying the Mardia test. If the multivariate normality criterion was not met, a polychoric (Likert scores) or tetrachoric (criterion scores) correlation matrix was constructed. Based on this matrix, an optimized parallel analysis was performed from the randomized generation of 500 submatrices (Timmerman, & Lorenzo-Seva, 2011). The suitability of the one-factor solution was studied with two indicators of suitability: Unidimensional Congruence (UniCo, appropriate when $>.5$) and the Explained Common Variance (ECV, appropriate when $>.85$). The accuracy of the one-factor solution was studied using the Determination Index (FDI, appropriate when $>.90$). Response bias was controlled for by accelerated Bootstrapping (BCa) and establishing a confidence interval for factor loadings on each item. The repeatability of the construct was studied using the generalized h-index (G-H). All the above estimators were used following the suggestions of Ferrando and Lorenzo-Seva (2017). Internal consistency was studied using the Greatest Lower Bound (GLB), McDonald's Omega (ω) and the standardized Cronbach's Alpha (α_s) (Trizano-Hermosilla, & Alvarado, 2016). Estimators of the model's goodness of fit to data used were: Root Mean Square Error of Approximation (RMSEA, which should be $<.07$), Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit Index (AGFI), Non-Normed Fit Index (NNFI) and Comparative Fit Index (CFI), all of them appropriate at $>.95$ (Hooper, Coughlan, & Mullen, 2008).

The FACTOR Program 10.8.03 was used for all these analyses (Ferrando, & Lorenzo-Seva, 2013). Sample subgroups were compared by analysis of variance and omega squared (ω^2) was used as the estimator of effect size, which is interpreted as small when $\omega^2=.01$, medium when $\omega^2=.06$ and large when $\omega^2=.14$. For comparison of categorical variables, the ji squared (χ^2) test and Cramer's V, which is interpreted as small when $V=.10$, medium when $V=.30$ and large when $V=.50$, were used. These analyses were performed using SPSS 22 statistical software (omega squared test was done manually based on the ANOVA provided by the program).

Results

The outlier detection study did not advise excluding any of the subjects from the sample (Mahalanobis Distance $p > .001$ in all cases).

Likert-type scoring

The descriptive statistics for the items are shown in Table 2.

Table 2. Descriptive statistics of the GHQ-12 items, Likert form.

Item	Mean	CI 95%	SD	Median	Variance	Skewness	Kurtosis	r_{it}
1	1.10	(1.06 - 1.14)	0.67	1	0.45	0.56	0.85	.39
2	0.82	(0.77 - 0.87)	0.89	1	0.79	0.77	-0.38	.47
3	1.08	(1.04 - 1.12)	0.70	1	0.50	0.82	1.26	.43
4	0.94	(0.91 - 0.98)	0.58	1	0.33	0.46	1.73	.41
5	1.29	(1.24 - 1.34)	0.93	1	0.87	0.20	-0.85	.58
6	0.94	(0.89 - 0.99)	0.87	1	0.75	0.60	-0.40	.63
7	1.03	(1.00 - 1.07)	0.67	1	0.45	0.58	0.97	.49
8	1.03	(0.99 - 1.06)	0.63	1	0.40	0.62	1.42	.53
9	0.85	(0.80 - 0.90)	0.92	1	0.85	0.79	-0.36	.68
10	0.67	(0.62 - 0.72)	0.88	0	0.77	1.10	0.20	.66
11	0.55	(0.50 - 0.60)	0.88	0	0.77	1.48	1.11	.61
12	0.97	(0.94 - 1.01)	0.68	1	0.46	0.70	1.24	.57

Note: r_{it} = Corrected item total correlation.

The Mardia test was applied to the Pearson correlations matrix, which showed that the distribution of items did not meet multivariate normality ($p < .05$), so a polychoric correlations matrix was constructed. Based on the new matrix, an optimized parallel analysis was made which suggested one-dimensionality of the scale with a single factor explaining 46% of accumulated variance (Table 3). The goodness-of-fit statistics for this one-factor solution were satisfactory: RMSEA=.07; NNFI=.97; CFI=.98; GFI=.98; AGFI=.98. The model repeatability index (Latent $h = .91$; Observed $h = .81$) was adequate, as was the suitability of the one-factor solution (UniCo = 0.97; ECV = 0.87). Internal consistency indicators were also appropriate: GLB=.93; $\omega = .89$; $\alpha_s = .89$. Table 4 shows the characteristics of the one-factor solution found. The results did not vary when the same tests were done separately in males and females.

Criterion Scoring

The descriptive statistics of the items are shown in Table 4. The Mardia test showed again that the distribution of the items did not meet multivariate normality ($p < .05$), so the tetrachoric correlations matrix was constructed. An optimized parallel analysis based on the new matrix suggested the unidimensionality of the scale with a single factor explaining 60.2% of the total variance of the test (Table 5). The

Table 3. One-factor solution (Likert scoring).

Item	Loads	Bca IC 95%	Communality
1	0.46	(0.40 - 0.51)	0.21
2	0.54	(0.49 - 0.58)	0.29
3	0.50	(0.45 - 0.55)	0.50
4	0.51	(0.46 - 0.56)	0.26
5	0.67	(0.63 - 0.70)	0.44
6	0.73	(0.70 - 0.76)	0.54
7	0.57	(0.52 - 0.61)	0.33
8	0.62	(0.57 - 0.66)	0.38
9	0.79	(0.76 - 0.82)	0.63
10	0.79	(0.76 - 0.82)	0.63
11	0.75	(0.72 - 0.79)	0.57
12	0.68	(0.65 - 0.71)	0.46

Note: Bca = Accelerated Bootstrap with bias scoring, 95% confidence interval.

goodness-of-fit statistics for this one-factor solution were satisfactory: RMSEA=.04; NNFI=.99; CFI=.99; GFI=.99; AGFI=.99. The model's Repeatability Index (Latent- $h = .95$; Observed- $h = .70$) and the suitability of the one-factor solution (UniCo=.99; ECV=.92) were also adequate. The internal consistency indicators were appropriate as well: GLB=.96; $\omega = .94$; $\alpha_s = .94$. Table 5 shows the characteristics of the one-factor solution found. The results did not vary when the same tests were given males and females separately.

Table 4. Descriptive statistics for the GHQ-12 items, criterion form.

Item	Mean	CI 95%	SD	Median	Variance	Skewness	Kurtosis	r_{it}
1	0.21	(0.19 - 0.23)	0.166	0	0.17	1.42	0.02	0.42
2	0.23	(0.20 - 0.25)	0.174	0	0.17	1.32	-0.26	0.45
3	0.18	(0.16 - 0.20)	0.149	0	0.15	1.65	0.73	0.50
4	0.11	(0.09 - 0.13)	0.099	0	0.10	2.48	4.14	0.49
5	0.40	(0.38 - 0.43)	0.241	0	0.24	0.39	-1.85	0.50
6	0.24	(0.22 - 0.27)	0.183	0	0.18	1.21	-0.55	0.57
7	0.18	(0.16 - 0.20)	0.147	0	0.15	1.67	0.80	0.51
8	0.16	(0.14 - 0.18)	0.134	0	0.13	1.87	1.48	0.60
9	0.23	(0.21 - 0.25)	0.178	0	0.18	1.27	-0.38	0.61
10	0.18	(0.16 - 0.20)	0.149	0	0.15	1.65	0.72	0.59
11	0.16	(0.13 - 0.17)	0.131	0	0.13	1.91	1.65	0.54
12	0.15	(0.13 - 0.17)	0.126	0	0.13	1.99	1.94	0.59

Note: r_{it} = Corrected item total correlation.

Table 5. One-factor solution (criterion scoring).

Item	Loads	Bca CI 95%	Comunality
1	0.61	(0.56 - 0.66)	0.38
2	0.64	(0.59 - 0.68)	0.41
3	0.72	(0.65 - 0.76)	0.51
4	0.75	(0.70 - 0.79)	0.56
5	0.71	(0.67 - 0.75)	0.51
6	0.77	(0.74 - 0.81)	0.60
7	0.73	(0.68 - 0.76)	0.53
8	0.84	(0.79 - 0.86)	0.70
9	0.83	(0.79 - 0.86)	0.69
10	0.81	(0.77 - 0.85)	0.66
11	0.77	(0.73 - 0.81)	0.59
12	0.83	(0.80 - 0.86)	0.69

Note: BCa= Accelerated Bootstrap with bias scoring, 95% confidence interval.

Sex differences

Table 6 shows the differences in answers to the GHQ-12 between males and females. Females said they had more problems concentrating, more depressive symptoms, less self-confidence and less perceived ability to cope with problems, but especially, more feelings of anxiety and tension than males, although all of this with a very small effect size.

Table 6. Gender differences in the responses to GHQ-12 items (Likert scoring).

	Males		Females		<i>F</i>	<i>p</i>	ω^2
	Mean	<i>SD</i>	Mean	<i>SD</i>			
Distraction	1.01	0.66	1.19	0.66	41.29	<.001	.02
Insomnia	0.72	0.86	0.92	0.91	29.53	<.001	.01
Emptiness	1.05	0.71	1.11	0.69	4.53	<.05	.01
Indecisiveness	0.91	0.58	0.98	0.58	7.26	<.01	.01
Stress	1.11	0.90	1.47	0.93	82.65	<.001	.04
Inability to work	0.83	0.85	1.05	0.87	33.11	<.001	.02
Enjoyment	1.02	0.67	1.05	0.66	1.52	.22	.01
Control	0.98	0.62	1.07	0.64	13.03	<.001	.01
Depression	0.73	0.89	0.98	0.94	39.11	<.001	.02
Self-confidence	0.56	0.84	0.78	0.90	33.89	<.001	.02
Inutility	0.44	0.80	0.66	0.94	33.75	<.001	.02
Happiness	0.94	0.66	1.01	0.70	5.58	<.05	.01

Note: ω^2 = Omega squared as an estimator of the effect size.

Scores

Following the author's instructions, the cutoff point should be three affirmative answers as an indicator of risk of poor mental health. This risk would be moderate for the mean plus the standard deviation (three, four or five affirmative answers) and severe for over six affirmative answers. Thus, the sample could be classified in three categories: no risk, moderate risk and severe risk (Table 7).

Another way of classifying the participants in our sample would be to use the Likert scores, with which scores not over the sample mean plus the standard deviation show no risk of poor mental health, between the mean plus one and plus two standard deviations show moderate risk, and over the mean plus two standard deviations show severe risk (Table 7).

Table 7. Distribution of the sample classified into three categories of risk of poor mental health.

	Males	CI95%	Females	CI95%	χ^2	V
<i>Criterion scores</i>						
No risk	69.9%	66.4 - 73.4	58.4%	54.7 - 62.1	34.23	.12
Moderate risk	17.2%	14.4 - 20.0	22.8%	19.6 - 26.0		
Severe risk	12.9%	10.4 - 15.4	18.8%	15.8 - 21.8		
<i>Likert scores</i>						
No risk	88.8%	86.8 - 90.6	82.1%	79.6 - 84.3	19.93	.10
Moderate risk	7.4%	5.9 - 9.2	11.7%	9.9 - 13.8		
Severe risk	3.8%	2.7 - 5.1	6.2%	4.8 - 7.8		

Discussion

The results of this study clearly show that the GHQ-12 should be considered a unidimensional test of psychological

distress, at least when applied to an adolescent population. Exploratory and semi-confirmatory (Lorenzo-Seva, & Ferrando, 2013) factor analysis were used, controlling for biases inherent in the item response method, in a randomized sam-

ple representative of the adolescent population (14-16-years-old) in a large city (Madrid). Unlike previous studies, usually using convenience sampling, obsolete methods inadequate for the nature of the data (Kaiser criterion, Scree-plot, Pearson correlation matrix, Cronbach's alpha, etc.), were avoided and more appropriate statistical indicators were used, both for the Likert-type response format and dichotomous scoring of the questionnaire. All of the statistics enable us to state that there is no doubt at all of the one-dimensionality of the test. These data are consistent with the proposals of the authors who attribute the appearance of two or three dimensions to the use of inadequate estimators, which are based on inadequate assumptions, such that when all those biases are controlled for the solution is always unidimensional (Hankins, 2008).

As a unidimensional test, the GHQ-12 is a reliable instrument, both overall and each item alone. What in other studies appear to be independent dimensions (social dysfunction, anxiety and loss of self-confidence) are simply components of a general construct of psychological distress, from which diagnostic approaches cannot be derived, but merely initial screening, which must be followed up with specific psychodiagnosis.

When gender differences were studied, females scored higher on many of the symptoms and the overall test score, as is usual. The real reasons for this tendency of females to report more distress, somatic and psychic symptoms is unknown (Davis, Matthews, & Twamley, 1999), a good number of possible reasons having been proposed: innate differences in somatic and visceral perception, differences in labeling, the description and reporting of symptoms, differences in education and socialization, sex differences in the incidence of abuse and violence, and gender prejudices in research as well as in clinical practice (Barsky, Peekna, & Borus, 2001). Whatever the reason, the data in our study sug-

gest that adolescent girls are more vulnerable than their male peers.

Applying the usual criterion scoring, we found that 30% of male and 42% of female adolescents were at risk of developing mental health problems. Doubtless, one of the main characteristics of this period of life is emotional instability derived from the change in roles, search for identity, the beginning of autonomy in decision-making, etc., which coincides with a neurologically critical period of synaptic pruning (Blakemore, & Choudhury, 2006). However, it seems exaggerated for the percentages mentioned to have any kind of clinical entity. In any case, a screening instrument like the GHQ-12 can enable the early detection of problems which often form the basis for establishment of psychopathological symptoms that persist through adulthood (Paus, Keshavan, & Giedd, 2008). The results found using the Likert scores (around 11% of males and 18% of females at risk of poor mental health) are closer to clinical reality, but it should not be forgotten that the GHQ-12 is a screening instrument and not meant for diagnosis.

In conclusion, the GHQ-12 is a unidimensional screening test for psychological distress, with excellent psychometric properties for its application in an adolescent population. Future studies should replicate these results in other populations (e.g., rural, adults, elderly, etc.), in other cultural settings and in clinical populations. The use of adequate statistical methods can enable old controversies to be overcome and favor proper application and interpretation of the results provided by the test.

Funding Sources: Funded by the Camilo José Cela University of Madrid, Spain. This research did not receive any specific funding from public, commercial, or nonprofit agencies.

Declarations of interest: none.

References

- Almeida, C. R., Coutinho, E. S. F., Silva, D. A., Oliveira, E. R. A., Bloch, K. V., & Viana, M. C. (2019). Exposure to breastfeeding and common mental disorders in adolescence. *Cadernos de Saude Publica*, 35(5), e00093718-e00093718. DOI: 10.1590/0102-311X00093718
- Baksheev, G. N., Robinson, J., Cosgrave, E. M., Baker, K., & Yung, A. R. (2011). Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students. *Psychiatry Research*, 187(1-2), 291-296. DOI: 10.1016/j.psychres.2010.10.010
- Bansal, V., Goyal, S., & Srivastava, K. (2009). Study of prevalence of depression in adolescent students of a public school. *Industrial Psychiatry Journal*, 18(1), 43-46. DOI: 10.4103/0972-6748.57859
- Barsky, A. J., Peekna, H. M., & Borus, J. F. (2001). Somatic symptom reporting in women and men. *Journal of General Internal Medicine*, 16(4), 266-275. DOI: 10.1046/j.1525-1497.2001.016004266.x
- Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, 47(3-4), 296-312. DOI: 10.1111/j.1469-7610.2006.01611.x
- Bowe, A. (2017). The cultural fairness of the 12-item General Health Questionnaire among diverse adolescents. *Psychological Assessment*, 29(1), 87-97. DOI: 10.1037/pas0000323
- Crutzen, R., & Peters, G. J. Y. (2017). Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, 11(3), 242-247. DOI: 10.1080/17437199.2015.1124240
- Davis, M. C., Matthews, K. A., & Twamley, E. W. (1999). Is life more difficult on Mars or Venus? A meta-analytic review of sex differences in major and minor life events. *Annals of Behavioral Medicine*, 21(1), 83-97. DOI: 10.1007/BF02895038
- Díaz-Olalla, J. M., & Benítez Robredo, T. (2015). *Estudio de Salud de la Ciudad de Madrid 2014*. Madrid: Madrid Salud, Ayuntamiento de Madrid. [Spanish]. http://www.madridsalud.es/publicaciones/OtrasPublicaciones/estudio_salud_madrid_2014.pdf
- Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona. Document available at: <http://psico.fccep.urv.cat/utilitats/factor/>
- Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, 78(5), 762-780. DOI: 10.1177/0013164417719308
- French, D. J., & Tait, R. J. (2004). Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *European*

- Child & Adolescent Psychiatry*, 13(1), 1-7. DOI: 10.1007/s00787-004-0345-7
- Glozah, F. N., & Pevalin, D. J. (2015). Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *Journal of Child & Adolescent Mental Health*, 27(1), 53-57. DOI: 10.2989/17280583.2015.1007867
- Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire*. London: Oxford University Press.
- Goldberg, D. P., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. Londres: Nfer-Nelson.
- Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing?. *Clinical Practice and Epidemiology in Mental Health*, 4(1), 10. DOI: 10.1186/1745-0179-4-10
- Hankins, M. (2008). The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health*, 8(1), 355. DOI: 10.1186/1471-2458-8-355
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Li, W. H., Chung, J. O., Chui, M. M., & Chan, P. S. (2009). Factorial structure of the Chinese version of the 12-item General Health Questionnaire in adolescents. *Journal of Clinical Nursing*, 18(23), 3253-3261. DOI: 10.1111/j.1365-2702.2009.02905.x
- Lobo, A., & Muñoz, P. E. (1996). *General Health Questionnaire. Guide for the use of the different validated versions in Spanish language* [Cuestionario de Salud General GHQ. Guía para el usuario de las distintas versiones en lengua española validadas]. Barcelona: Masson.
- Lopez-Castedo, A., & Fernandez, L. (2005). Psychometric properties of the Spanish version of the 12-item General Health Questionnaire in adolescents. *Perceptual and Motor Skills*, 100(3), 676-680. DOI: 10.2466/pms.100.3.676-680
- Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2 A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, 37(6), 497-498. DOI: 10.1177/0146621613487794
- Makowska, Z., Merecz, D., Moscicka, A., & Kolasa, W. (2002). The validity of general health questionnaires, GHQ-12 and GHQ-28, in mental health studies of working people. *International Journal of Occupational Medicine and Environmental Health*, 15(4), 353-362.
- Mann, R. E., Paglia-Boak, A., Adlaf, E. M., Beitchman, J., Wolfe, D., Wekerle, C., ... & Rehm, J. (2011). Estimating the prevalence of anxiety and mood disorders in an adolescent general population: An evaluation of the GHQ12. *International Journal of Mental Health and Addiction*, 9(4), 410-420. DOI: 10.1007/s11469-011-9334-5
- Ogawa, S., Kitagawa, Y., Fukushima, M., Yonehara, H., Nishida, A., Togo, F., & Sasaki, T. (2019). Interactive effect of sleep duration and physical activity on anxiety/depression in adolescents. *Psychiatry Research*, 273, 456-460. DOI: 10.1016/j.psychres.2018.12.085
- Ojio, Y., Nishida, A., Shimodera, S., Togo, F., & Sasaki, T. (2016). Sleep duration associated with the lowest risk of depression/anxiety in adolescents. *Sleep*, 39(8), 1555-1562. DOI: 10.5665/sleep.6020
- Padrón, A., Galán, I., Durbán, M., Gandarillas, A., & Rodríguez-Artalejo, F. (2012). Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Quality of Life Research*, 21(7), 1291-1298. DOI: 10.1007/s11136-011-0038-x
- Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence?. *Nature Reviews Neuroscience*, 9(12), 947-957. DOI: 10.1038/nrn2513
- Pedrero-Pérez, E. J., Mora-Rodríguez, C., Rodríguez-Gómez, R., Benítez-Robredo, M. T., Franco, A. O., González-Robledo, L., & Méndez-Gago, S. (2019). Síntomas prefrontales asociados al uso problemático de las Tecnologías de la Información y la Comunicación (TIC) en adolescentes. *Behavioral Psychology/Psicología Conductual*, 27(2), 257-273.
- Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., & Ponsoda, V. (2014). The impact of ambiguous response categories on the factor structure of the GHQ-12. *Psychological Assessment*, 26(3), 1021-1030. DOI: 10.1037/a0036468
- Rommel, M., Braehler, E., Roth, M., & Glaesmer, H. (2013). What is the General Health Questionnaire-12 assessing?. *Comprehensive Psychiatry*, 54(4), 406-413. DOI: 10.1016/j.comppsy.2012.10.010
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, 24(2), 282-292. doi:10.1037/a0025697
- Sánchez-López, M. D. P., & Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population. *Psicobema*, 20(4), 839-843.
- Sarkova, M., Nagyová, I., Katreniaková, Z., Geckova, A. M., Orosová, O., Middel, B., ... & van den Heuvel, W. (2006). Psychometric evaluation of the General Health Questionnaire-12 and Rosenberg Self-esteem Scale in Hungarian and Slovak early adolescents. *Studia Psychologica*, 48(1), 69-79.
- Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological Assessment*, 17(2), 231-236. DOI: 10.1037/1040-3590.17.2.231
- Sheldrick, R. C., Bennayan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Thresholds and accuracy in screening tools for early detection of psychopathology. *Journal of Child Psychology and Psychiatry*, 56(9), 936-948. DOI: 10.1111/jcpp.12442
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. DOI: 10.1037/a0023353
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, 769. DOI: 10.3389/fpsyg.2016.00769
- Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century?. *Journal of Psychoeducational Assessment*, 29(4), 377-392. DOI: 10.1177/0734282911406668