



Assessment of the factor structure and reliability of the 28 item version of the General Health Questionnaire (GHQ-28) in El Salvador

Pat Gibbons¹, Hilda Flores de Arévalo, and Mauricio Mónico
(Universidad Tecnológica de El Salvador, El Salvador)

(Recibido 8 enero 2003 / Received January 8, 2003)
(Aceptado 6 noviembre 2003 / Accepted November 18, 2003)

ABSTRACT. We wished to assess both the factor structure and the reliability of the 28-item version of the General Health Questionnaire (GHQ-28) in a non-clinical population in El Salvador. The GHQ-28 was applied to a large sample ($n = 714$) of university students in El Salvador. A principal components analysis was applied and a four-factor solution selected. A sub-group of the participants were retested after an average of 3 months to assess retest reliability while internal consistency was assessed using the Cronbach's alpha test. The factor analysis results corresponded closely to the original factor structure, though with a relatively higher inter-correlation found between the resulting scales, especially the 'anxiety and insomnia' and 'somatization' scales. The test-retest and internal consistency measures reached values .70 or higher in each case. The findings indicate a remarkable consistency in the factor structure of the GHQ-28 with result from other cultural settings, supporting Goldberg's hypothesis of a common language of psychological distress between cultures. The 'severe depression' and 'social dysfunction' scales appear to be relatively stable and independent while the 'anxiety and insomnia' and 'somatization' scales are more highly correlated. The reliability of the instrument in El Salvador is satisfactory.

KEYWORDS. General Health Questionnaire (GHQ-28). University sample. Factor analysis. Reliability Instrumental study.

¹ Correspondence: Department of Psychiatry. Naas General Hospital. Naas. Co. Kildare (Republic of Ireland).
E-mail: patgibbons@yahoo.com

RESUMEN. Quisimos evaluar tanto la estructura factorial como la confiabilidad de la versión de 28 ítem del Cuestionario de Salud General (GHQ-28) en una muestra no-clínica en El Salvador. Aplicamos el GHQ-28 a una muestra de 714 de alumnos universitarios en El Salvador. Utilizamos un análisis de componentes principales y seleccionamos una solución de 4 factores. Reaplicamos la prueba a un grupo de los participantes después de un promedio de 3 meses, para evaluar la confiabilidad de 'retest' del instrumento y calculamos la confiabilidad interna del GHQ-28 a través de la prueba alfa de Cronbach. Los resultados del análisis factorial correspondieron estrechamente con la estructura factorial antes identificada, aunque encontramos una correlación más alta entre las escalas, sobre todo entre Ansiedad e insomnio y Somatización. La fiabilidad test-retest y la consistencia interna alcanzaron valores de 0,70 o superiores. Los resultados del análisis factorial indican una gran coincidencia con las conclusiones de otros entornos culturales, lo que apoya la hipótesis de Goldberg de que hay una manera común de comunicar angustia psicológica en culturas distintas. Las escalas de Depresión severa y de Disfunción social parecen ser relativamente estables e independientes, mientras que las de Ansiedad e insomnio y de Somatización se correlacionan más. La fiabilidad del instrumento es satisfactoria.

PALABRAS CLAVE. Cuestionario de Salud General (GHQ-28). Muestra universitaria. Análisis factorial. Confiabilidad. Estudio instrumental.

RESUMO. Quisemos avaliar tanto a estrutura factorial como a fiabilidade da versão de 28 itens do Questionário de Saúde Geral (GHQ-28) numa amostra clínica em El Salvador. Aplicámos o GHQ-28 a uma amostra de 714 de alunos universitários em El Salvador. Utilizámos uma análise de componentes principais e seleccionámos uma solução de 4 factores. Reaplicámos a prova a um grupo de participantes de pois de uma média de 3 meses, para avaliar a fidelidade de reteste do instrumento e calculámos a fidelidade interna do GHQ-28 através da prova alpha de Cronbach. Os resultados da análise factorial corresponderam estreitamente à estrutura factorial anteriormente identificada, ainda que tivéssemos encontrado uma correlação mais alta entre as escalas, sobretudo entre Ansiedade e insónia e Somatização. A fidelidade teste-reteste e a consistência interna alcançaram valores de 0,70 ou superiores. Os resultados da análise factorial indicam uma grande coincidência com as conclusões de outros meios culturais, o que apoia a hipótese de Goldberg de que há uma forma comum de comunicar angústia psicológica em culturas distintas. As escalas de Depressão severa e de Disfunção social parecem ser relativamente estáveis e independentes, enquanto que as de Ansiedade e insónia e de Somatização se correlacionam mais. A fidelidade do instrumento é satisfatória.

PALABRAS CHAVE. Questionário de Saúde Geral (GHQ-28). Amostra universitária. Análise factorial. Fidelidade. Estudo instrumental.

Introduction

The General Health Questionnaire (GHQ) was originally described by Goldberg (1972) as a self-administered screening instrument for psychiatric disorder in non-clinical populations, and has come to be is the most widely used of such instruments

worldwide (Werneke, Goldberg, Yalcin, and Ustun, 2000). It has gained widespread acceptance for its brevity, ease of administration and acceptability to the user. The instrument examines functioning in two main areas, firstly of one's ability to carry out one's usual healthy activities, and secondly of the recent development of subjective symptoms of psychological distress (Goldberg and Williams, 1988). A number of different versions of the instrument exist, ranging in size from 12 to 60 items, each of which is scored on a 4 point Likert-type scale of severity or alternatively according to a 0-0-1-1 system described by the authors as being the most useful in identifying psychiatric 'caseness'. The 28-item version of the GHQ is the only version that provides subscale measures of more specific domains of psychopathology (Goldberg and Hillier, 1979). As well as the global score, there are four subscales consisting of 7 items in each case, which are labelled Severe depression, Anxiety and insomnia, Somatization and Social dysfunction. The item content of these subscales was selected so as to maximise the differentiation of anxiety from depression. Validation of the subscales against a structured clinical interview has revealed that the Anxiety and insomnia and the Severe depression subscales both correlated equally with interview ratings of anxiety and depression, though the Somatization subscale correlated less well with interview rating of the same domain (Goldberg and Williams, 1988). The GHQ has been translated into 38 languages, and has been studied in a wide range of cultural settings, described in over 700 articles currently listed in the 'Medline' database of medical journals maintained by the National Library of Medicine of the United States. Factor analysis of GHQ-28 data from 15 different cultural and linguistic settings worldwide has confirmed the stability of the Severe depression and Social dysfunction subscales, though indicating significant overlap of the Anxiety and insomnia and Somatic symptoms domains (Werneke *et al.*, 2000). The Spanish language version of the GHQ has received relatively little research attention outside of Spain and is not cited in any previous study from Central America, despite the fact that the overwhelming majority of Spanish speakers live in Latin America and the USA.

We wished to assess the factor structure of the GHQ-28 in a Salvadorean university student sample and to compare the emerging factors with those identified in previous studies. We further wished to assess the test-retest reliability and internal consistency of the instrument in El Salvador.

The structure of this article follows the procedures suggested by Bobenrieth (2002) for research papers on health science, and the general methodology for classification and description proposed by Montero and León (2002).

Method

Sample

The study sample was identified using a non-probabilistic quota sampling method, so as to be representative of the overall student population of the four major faculties of the Technological University of El Salvador. This is the second largest university in the country, is located in the centre of San Salvador and draws its student body from a broad range of socio-economic groups. Of the total sample of 732 students selected

to take part in the study, 714 (97.5%) participated. Four subjects failed to complete four or more of the original items, and these cases were excluded from the subsequent analysis. Of the remaining cases, where isolated items were missed, no specific pattern emerged to suggest that these were not random errors. These latter cases were included, replacing the missing values with the mean score for that item. The mean age of the sample was 22.4 years, and 52% were male.

Procedures

The Spanish language version of the GHQ-28 (Muñoz, Vázquez, Pastrana, Rodríguez, and Oneca, 1978) was administered to students from the various faculties during a psychology class that is part of the general university curriculum. The questionnaire was administered on a group basis along with a number of other self-report tests as part of a larger study of the psychological effects of childhood trauma. Forty-five participants repeated the GHQ after a lapse of 2-6 months to provide an estimate of test-retest reliability. Likert scale data for each of the 28 items was entered into a principal components analysis. A four-factor solution was selected on the pragmatic basis that it would facilitate comparison with existing studies. Although an oblique rotation has been recommended when the emerging factors are expected to be significantly correlated (Elton, Patton, Weyerer, Diallina, and Fichter, 1988), we applied a varimax rotation, which provides results which are easier to interpret (Werneke *et al.*, 2000) and is the technique used in previous studies. A loading score of .40 was used as the cut-off score for assigning items to a factor, and items were assigned to the factor to which they revealed the highest loading. Test-retest analysis was assessed by analysing the data as both categorical (i.e. comparing 'caseness' between the two measures using the kappa statistic), and as ratio scale, correlating the mean Likert scale scores between the first and second administration of the test (Pearson's 'r'). Internal consistency of the overall and subscale scores was assessed using the Cronbach's 'alpha' statistic. All statistical analysis was carried out using the version 9.0 of the SPSS software.

Results

As expected, the mean GHQ score (Likert method) was higher for females (mean = 26) than for males (mean = 20; $t = -5.6$, $df = 704$, $p < .001$).

Factor analysis

A principal components analysis ($n = 710$) was carried out, selecting a four-factor solution and using a varimax rotation. These four factors were found to account for 54% of the overall variance, and as can be seen from Table 1, all the items load onto their original subscales with loading values of .40 or greater in each case. Three items load equally onto one other factor as well as their original scale. Items A7 ("Have you been having hot and cold spells?") and B3 ("Have you been feeling constantly under strain?") both load equally onto the 'somatization' as well as 'anxiety and insomnia' factors, while item D5 ("Have you found at times that you couldn't do anything because your nerves were too bad?") loads equally onto the Severe depression as well as the

Anxiety and insomnia factor, as also reported by Goldberg and Williams (1988). In addition, items A1, A5 and A6 of the Somatization scale, B7 of the Anxiety scale, C5 of the Social dysfunction scale and D2 of the Depression scale all load to a lesser but significant degree on to at least one other factor. The fact that 9 of the 28 items of the GHQ-28 load significantly onto two factors indicates that the underlying factors are significantly correlated. Nevertheless, the mean factor loadings of the 7 items for each of the subscales is very much greater for their original subscale than for any of the other 3 subscales (Table 2).

The high degree of overlap in the domains indicated by these factor loadings, especially between Anxiety and insomnia and Somatization is confirmed by estimating the Pearson's product-moment correlations between the scales, indicated in Table 3.

TABLE 1. Factor loadings of the items of the GHQ-28
(following the coding of the original version in english).

<i>Somatization</i>		<i>Anxiety and insomnia</i>		<i>Social dysfunction</i>		<i>Severe Depression</i>	
A1	.60	B1	.80	C1	.60	D1	.60
A2	.70	B2	.70	C2	.40	D2	.60
A3	.70	B3	.50	C3	.70	D3	.70
A4	.70	B4	.60	C4	.70	D4	.80
A5	.50	B5	.50	C5	.60	D5	.50
A6	.60	B6	.50	C6	.60	D6	.80
A7	.40	B7	.50	C7	.50	D7	.70
Mean	.59	Mean	.57	Mean	.59	Mean	.67

TABLE 2. Mean factor loadings of the individual items
of the four scales on the four factors.

	<i>Somatization</i>	<i>Anxiety and insomnia</i>	<i>Social dysfunction</i>	<i>Severe depression</i>
<i>Somatization</i>	.67	.41	.26	.27
<i>Anxiety and insomnia</i>	.43	.67	.38	.41
<i>Social dysfunction</i>	.29	.23	.63	.34
<i>Severe depression</i>	.32	.34	.41	.73

	Anxiety and insomnia	Social dysfunction	Severe depression
Somatization	.70	.51	.50
Anxiety and insomnia	1	.62	.65
Social dysfunction	-	1	.61

TABLE 3. Inter-correlations of the GHQ-28 subscales.Note: $p < 0.001$ in each case.*Reliability analysis*

The test-retest reliability of the GHQ-28 can be assessed analysing the data as either categorical or dimensional in nature, as discussed in the introduction (i.e. either by using a cut-off score to define 'cases' and 'non-cases', or alternatively by analysing the GHQ score as a continuous ratio-scale variable). For analysis as a categorical measure of 'caseness', we used the 6/7 cut-off point identified as maximising the efficiency (sensitivity and specificity) of the Spanish language version of the GHQ-28 in Spain (Lobo, Perez-Echevarría, and Artal, 1986) and Chile (Goldberg *et al.*, 1997). This cut-off point gives a prevalence of 'caseness' among the overall sample of 35%. For the test-retest sample ($n = 42$), the value of kappa was .74 ($p < .001$), giving an 88% correspondence between the two applications of the instrument (Table 4). Analysing the Likert scale data for each item, we estimated the Pearson's 'product moment' correlation coefficient for the global and subscale scores (Table 5). Internal consistency of the global and subscale scores obtained from the total sample was assessed by estimating the value of Cronbach's alpha (Table 5).

TABLE 4 The test-retest reliability of the GHQ-28 (categorical measure)

SECOND APPLICATION OF THE GHQ	FIRST APPLICATION OF THE GHQ-28		
	Non-case	Case	Total
Non-case	25	1	26
Case	4	12	16
Total	29	13	42

SCALE	Value of Pearson's γ	Cronbach's α
Global value	.85	.86
Somatization	.74	.81
Anxiety and insomnia	.87	.86
Severe depression	.71	.87
Social dysfunction	.67	.79

TABLE 5. Test-retest reliability and internal consistency of the GHQ-28.

Note: $p < .001$ in each case.

Discussion and conclusion

The current investigation addresses two principal issues in relation to the GHQ-28. The first of these is whether the component structure of the instrument in El Salvador supports Goldberg's proposition that there is a common language of psychological distress that is independent of the cultural context. Secondly, we wished to assess the reliability of the instrument in El Salvador. The factor structure originally described for the GHQ-28 was derived from data gathered using the GHQ-60 in Philadelphia in the U.S.A. and in Croydon and Manchester in the U.K. by Goldberg and Hillier (1979). Since then, replication studies have been carried out in a broad range of cultural settings, including at least one Latin American sample from Chile, quoted in a review of 15 centres in a WHO study of psychological disorders in a general health care setting (Werneke *et al.*, 2000). This international study found that the C (Social dysfunction) and the D (Severe depression) factors were remarkably stable between centres and were consistent with the original factor structure of Goldberg and Hillier, with item D5 being the only item repeatedly loading outside of the original scale. However, the A scale (somatization) and B scale (anxiety and insomnia) items loaded onto both the A and B factors, indicating a high degree of correlation between these factors. Four out of 6 ethnically diverse samples of 15 year old schoolgirls from Britain, Greece and Turkey reviewed by Elton *et al.* (1988) reveal factor structures similar to the original scales, when analysed individually and when grouped into a single large sample. Again, D5 was found to load onto the Anxiety and insomnia scale, while D1 and A1 loaded onto the Social dysfunction scale. The factors that emerge from our El Salvadorean sample are remarkably consistent with these previous studies, in that the depression and social dysfunction scales emerge as being relatively independent (apart from the D5 item) while the anxiety and somatization factors are more highly inter-correlated. In fact, there appears to be a greater degree of cross-loading of items between factors in our study than that reported in other samples and the average inter-scale correlation (.61) is also somewhat higher than that reported by Banks (1983) [.47] and Lobo *et al.* (1986) [.51]. These findings suggest that the GHQ factors are less clearly differentiated in the El Salvadorean sample than in previous studies.

Despite the high correlations between the subscales, the consistency of the factor structure with that reported from previous studies in different cultural settings indicates that the depression and social dysfunction factors represent distinct and valid domains of psychopathology, though the evidence for the existence of distinct anxiety and somatization domains is equivocal, as previously suggested by Werneke *et al.* (2000). The evident consistency of the factor structure is matched by adequate reliability scores in our sample. The internal consistency coefficients are satisfactory, and the value for the global scale corresponds closely to the results quoted by Goldberg and Williams (1988), who reported a mean value for Cronbach's alpha of .85 for the GHQ-12 and .93 for the GHQ-60 (values for the GHQ-28 subscales were not quoted). Thus the homogeneity of the items of the instrument appears to be a relatively stable quality whatever the cultural context in which it is applied. The test-retest coefficients are quite high given the considerable lapse of time (range: 2-6 months, mean approximately 3 months)

between the two applications of the test, and the fact that the test is designed to measure relatively acute changes in symptoms, which would be expected to fluctuate significantly over this period of time. In previous studies, the test-retest correlations for non-clinical groups have ranged from 0.36 in a sample of medical students retested after two years (Firth-Cozens, 1987) to .58 in a sample of school leavers retested after 6 months (Layton, 1986). The test-retest correlations for clinical samples have tended to be higher, for example .85 in a sample of neurological patients retested after 5-7 days (Depaulo and Folstein, 1978) and .90 in a sample of stroke patients retested after 8 months (Robinson and Price, 1982). The high test-retest correlations in our study correspond more closely to a clinical than a non-clinical sample and may reflect the relatively high prevalence of 'caseness' in our study population.

The high 'case' prevalence in our sample is consistent with previous studies using the Spanish version of the GHQ. The threshold for case definition for Latin American samples has been reported to be generally higher than in other settings (Lewis and Araya, 1995). Similarly, in a recent WHO study using the 12-item version of the GHQ in 15 primary care settings worldwide, the highest average scores were found in Santiago de Chile (Goldberg, Oldehinkel, and Ormel, 1998). It should be unsurprising that the prevalence rate of 'caseness' in the current study is higher than that reported from most samples in Western Europe, despite the higher threshold used, but is very similar to that reported for samples from Spain. Herran *et al.* (1999) reported a case prevalence of 33% and Jiménez-Cruzado, González-Botella, Penaroja-González, Nolazco-Bonmatí, Pérez-Navarro, and De la Hoz Roja (1993) of 35%, while Medina-Mori *et al.* (1985) reported a case prevalence of 34% in Mexico. Nevertheless, these latter studies were undertaken with samples drawn from primary care populations in which a higher average GHQ score might be expected than in non-clinical samples (as was used in the current study), due to an increased prevalence of somatic symptoms and associated psychological distress. The extent to which the high case prevalence in this university sample represents a genuinely higher prevalence of psychopathology or alternatively a culturally determined willingness to report symptoms of psychological distress is unclear at this stage. A rigorous external validation study, using a structured diagnostic interview technique, is required to clarify this issue.

The GHQ-28 is revealed to be a robust screening instrument for minor psychopathology, with a remarkably stable factor structure when applied to an El Salvadorean non-clinical sample when compared to previous findings from diverse cultural settings, indicating that the factors have a validity that does indeed cross cultural boundaries and supporting Goldberg's hypothesis of a common language of psychological distress in different cultures. The high case prevalence reflects previous results from Latin America and Spain, and may contribute to the relatively high inter-correlation between factors that is found in our sample. The test-retest reliability is somewhat higher than in previous studies using non-clinical samples, and we suggest that the coefficients are minimum measures of reliability, given that some of the variance in the results between applications can be explained by genuine variation in the level of symptoms. The internal consistency of the items in the global scale and the subscales is likewise shown to be stable and highly satisfactory. In summary, our data support the cross-cultural validity of the GHQ and reinforce its position as the most widely used

screening instrument for general psychopathology.

References

- Banks, M.H. (1983). Validation of the General Health Questionnaire in a young community sample. *Psychological Medicine*, 13, 349-353.
- Bobenrieth, M. (2002). Normas para la revisión de artículos originales en Ciencias de la Salud. *Revista Internacional de Psicología Clínica y de la Salud/International Journal of Clinical and Health Psychology*, 2, 509-523.
- Depaulo, J.R. and Folstein, M.F. (1978). Psychiatric disturbances in neurological patients: Detection, recognition and hospital care. *Annals of Neurology*, 4, 225-228.
- Elton, M., Patton, G., Weyerer, S., Diallina, M., and Fichter, M. (1988). A comparative investigation of the principal structure of the 28-item version of the General Health Questionnaire (GHQ). *Acta Psychiatrica Scandinavica*, 77, 124-132.
- Firth-Cozens, J. (1987). Emotional distress in junior house officers. *British Medical Journal*, 295, 533-536.
- Goldberg, P. (1972). *The detection of Psychiatric Illness by Questionnaire. Maudsley Monograph No. 21*. Oxford: Oxford University Press.
- Goldberg, D.P., Gater, R., Sartorius, N., Ustun, T.B., Piccinelli, M., Gureje, O., and Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, 27, 191-197.
- Goldberg, P. and Hillier, V.F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, 9, 139-145.
- Goldberg, D.P., Oldehinkel, T., and Ormel, J. (1998). Why the GHQ threshold varies from one place to another. *Psychological Medicine*, 28, 915-921.
- Goldberg, P. and Williams, P. (1988). *A User's Guide to the General Health Questionnaire*. Windsor: NFER-NELSON.
- Herrán, A., Vázquez-Barquero, J.L., Artal, J., García, J., Iglesias, C., Montejo, J., and Diez-Manrique, J.F. (1999). The recognition of mental disease in primary health care and its determining factors. *Actas Españolas de Psiquiatría*, 27, 87-95.
- Jiménez-Cruzado, L., González-Botella, A., Peñaroja-González, D., Nolazco-Bonmatí, A., Pérez-Navarro, I., and De la Hoz Roja, J. (1993). Factors associated with the prevalence of psychiatric morbidity in a health area. *Atención Primaria*, 11, 465-469.
- Layton, C. (1986). Note on test re-test characteristics of the General Health Questionnaire. *Perceptual and Motor Skills*, 62, 221-222.
- Lewis, G. and Araya, R.I. (1995). Is the General Health Questionnaire (12 item) a culturally biased measure of psychiatric disorder? *Social Psychiatry and Psychiatric Epidemiology*, 30, 20-25.
- Lobo, A., Pérez-Echevarría, M.J., and Artal, J. (1986). Validity of the scaled version of the General Health Questionnaire (GHQ-28) in a Spanish population. *Psychological Medicine*, 16, 135-140.
- Medina-Mora, M.E., Padilla, G.P., Mas, C.C., Ezban, B.M., Caraveo, A.J., Campillo, S.C., and Corona, J. (1985). The prevalence of mental disorders and risk factors in a population of a general medical practice. *Acta Psiquiátrica y Psicológica de América Latina*, 31, 53-61.
- Montero, I. and León, O.G. (2002). Clasificación y descripción de las metodologías de investigación en Psicología. *Revista Internacional de Psicología Clínica y de la Salud/International Journal of Clinical and Health Psychology*, 2, 503-508.
- Muñoz, P.E., Vázquez, J.L., Pastrana, E., Rodríguez, F., and Oneca, C. (1978). Study of the

- validity of Goldberg's 60-item GHQ in its Spanish version. *Social Psychiatry*, 13, 99-104.
Robinson, R.G. and Price, T.E. (1982). Post-stroke depressive disorders: a follow-up study of 103 patients. *Stroke*, 13, 635-641.
Werneke, U., Goldberg, P., Yalcin, I., and Ustun, B.T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine*, 30, 823-829.