

Fiji norms and a cross-cultural clinical validation of the general health questionnaire (GHQ 20)

Abstract: Norms are presented for the 20-item version of the General Health Questionnaire (GHQ 20) that were obtained from a non-clinical group of the Fiji population. The data was required to give meaning to the results that had been obtained from a group of former hostages and the families of hostages that were still in custody. With the omission of two items, the outcome endorsed the robust psychometric properties of the instrument in terms of reliability and factor structure, and supported its use with quite a different cultural population from that for which originally it was designed. A validation study also showed expected differences between the normative group and the hostage group, with members of the hostage group classified after clinical interviews showing the most significant stress and receiving the highest scores on all four subscales as well as the full scale. (Pacific Health Dialog 2003, Vol. 10 (2); Pg 62-65)

A.J.W. Taylor*
H.S. Aghanwa**
F.H. Walkey*

Introduction

The accurate documentation of symptoms of mental trauma is a prerequisite for their confirmation by clinical interview. In the process difficulties abound, especially when attempts are made to transfer the diagnoses that originated in the Western world to other cultures and vice versa (DSM IV TR 2000, Appendix 1: Young, 2001, ch. 8). With regard to the diagnosis of posttraumatic stress disorder (PTSD) alone De Girolamo and McFarlane (1996) point out that 'the differences between cultural patterns, social structures, and coping behaviours of developed and developing countries may significantly influence the incidence, severity, and psychosocial outcome'. They go on to complain that few psychiatric studies have actually been carried out in developing countries.

The present authors were obliged to consider such cultural differences when making a clinical appraisal of a sample of hostages and their families in Fiji after the parliamentary coup that took place there in May 2000 (Taylor, Nailatikau, & Walkey, 2002). Subsequently they set out to obtain norms for the GHQ (Goldberg, 1972; 1978; Goldberg & Williams, 1988) from a normative sample of the community in Fiji, and to consider the questionnaire's psychometric adequacy. They chose the GHQ because it a) was designed for screening a normal population for self-reported symptoms of psychological disorder, b) had well-established psychometric

properties, and c) had been used in other parts of the world in many translations - although in the present instance the English language version was used because the participants all had English as their second language. Although versions of different length were available that also could be scored by different methods, the researchers preferred the GHQ 20-item (Siegert, McCormick, Taylor, & Walkey, 19874) to others because it was brief and had been shown to retain four of the factors of longer versions. They adopted the 'simple Likert' method for the participants to recording their responses, because it gave a measure of intensity to supplement that of sensitivity.

Method

The sample consisted of 166 medical students and staff members of the School of Medicine and the Colonial War Memorial Hospital in Suva. None was a clinical patient or regarded as being in need of psychological treatment. None was directly related to the hostages and their families, but like members of the general community all were somewhat affected by the prolonged economic and social repercussions of the political coup that had taken place in their country (Sharma, 2001).

The data was processed with the use of the SPSS10.0.0 Package (9) with confirmatory factor analyses undertaken using Amos 4.0 (Arbuckle & Wothke (1999).

Initially a single component analysis was undertaken to identify items that in Fiji might evoke responses that were not suggestive of psychological dysfunction (as revealed by responses to the measure overall). As a result, items #2 and #18 were found to be unrelated to the responses that the population made to the remaining items, and they were excluded from subsequent analyses.

Confirmatory analyses were made of the four-factor structure and of the single general factor, for the purpose

* School of Psychology, Victoria University of Wellington, PO Box 600, Wellington, NZ; ** Department of Psychiatry Fiji School of Medicine, Suva, Fiji-Islands.

of seeking confirmation that the actual response data observed was not unacceptably different from that which could be expected from the theoretical model. Each analysis employed four of the highly correlated indicators of the 'goodness of fit' to examine the relationship between the response data and the model under scrutiny. In this procedure it should be noted that:

1. The Chi Square (X^2) was used to evaluate the significance of differences between the proposed model and the observed data. But contrary to the manner in which normally the results of X^2 analysis are sought, in the present instance to be acceptable the differences needed to be small, and therefore not statistically significant.
2. The convention was followed that the Chi Square to degrees of freedom ratio ($X^2:df$) should be less than 2.00, because many sets of apparently very good data produce a significant chi square value. Yet there is no specific statistical test to determine the significance either of this ratio, or for the two following indices of fit.
3. The Goodness of Fit index (GFI) should approach 1.00, although 0.95 is generally regarded as adequate, with the term 'marginally adequate' being used sometimes in cases where this criterion is not met.
4. The Root Mean Square of Approximation (RMSEA) is an index of the inadequacy of the model. In practice it should be close to zero, although 0.08 or less meets with general acceptance.
5. To minimise the effects of error variance in these analyses, two factor-parcels of items for each factor were examined in the four-factor model, and four domain parcels of items were examined in the one factor model (Kishton & Widaman, 1994)).

The confirmatory factor analyses were followed by calculations of the means, standard deviations, and estimates of the reliability for each subscale and for the full scale. Finally, a table of norms was created from the responses of this community sample.

A clinical classification of the members of the Hostage Group at the time of the crisis provided an opportunity to make an assessment of the validity of the GHQ(20) as an indicator of stress. A total of 41 members of the

Hostage Group and their families were interviewed at the time of the hostage crisis, and were diagnosed as either symptomatic (having symptoms of Post Traumatic Stress) or as non-symptomatic. Then the mean scores of those groups were compared with those of the normative group both on the subscales and on the full GHQ(20).

Results

a) Confirmatory Factor Analyses

1. With 14 degrees of freedom, the four factor analysis gave a X^2 value of 22.99 ($p < .10$) and a $X^2:df$ ratio of 1.64. The GFI was 0.97, and the RMSEA=0.06.
2. With 2 degrees of freedom, the single factor analysis gave a chi square value of 0.28 ($p < .90$) and a $X^2:df$ ratio of 0.14. The GFI was 1.00, and the RMSEA 0.00.

b) Reliability Estimates, Summary Statistics, and Norms

1. The reliability estimates, means, standard deviations, and norms are given for the four subscales and for the full (18 item) scale in Table 1.

Table 1. Showing Norms, Summary Statistics and Reliabilities of the Four Sub-scales and of the Eighteen Selected Items of the Full GHQ20

Percentiles	Raw Scores*				
	GI	SP	AX	SD	Total
1-10	0-2	1	0	0	0-4
11 - 20	2	1	0	0	5
21-30	3	2	1	0	6-7
31-40	3	3	1	0	8
41-50	3	4	2	0	9-10
51-60	3	5	3	0	11
61-70	4	6	4	0	12-14
71-80	4	7	5	1	15
81-90	5	8	6	2	16-19
91-95	6-7	9	7	3-4	20-22
96-100	8-12	10-15	8-15	5-12	23-54
Mean	3.44	4.08	2.63	0.55	10.65
Std. Dev.	1.62	2.90	2.44	1.34	6.16
Alpha	.60	.79	.79	.79	.85
Split Half	.66	.77	.77	.66	.66

*Note: GI = General Illness (Items #1, #3, #4, #5) SP = Sleep Problems (Items #6 - #10) AX = Anxiety (Items #11 - #15) SD = Severe Depression (Items #16, #17, #19, #20)

c) Comparison of normative and clinically classified hostage groups.

As a broad measure of the validity of the GHQ(20), the normative and clinically classified hostage groups were compared. As was expected, the hostage group indicated higher levels of disturbance on all four subscales and on the full scale. However, the focus of the validation study was also on the capacity of the GHQ(20) to reflect the outcomes of the clinical

assessment and differential classification within the hostage group. The scores of the normative group, the hostage group as a whole, and of the members of the hostage group clinically diagnosed as either symptomatic or non-symptomatic, are shown in Table 2.

Confirmatory Factor analyses

With the two items excluded from consideration, all four of the indices of goodness of fit used in the first factor analysis confirmed the expected four-factor

Table 2. Summaries of the test scores of the normative group, the Hostage Group as a whole, and of the two hostage groups identified as either acutely stressed or relatively unstressed during the hostage crisis

Scale	Group	Mean	Standard Deviation	
General Illness	Normative	2.98	0.89	161
	Hostage	6.00	2.85	41
	Non-symptomatic	3.46	1.90	13
	Symptomatic	7.18	2.42	28
Sleep Problems	Normative	2.95	1.65	164
	Hostage	9.29	5.03	41
	Non-symptomatic	4.46	5.50	13
	Symptomatic	11.54	2.76	28
Anxiety	Normative	2.16	1.72	165
	Hostage	8.00	4.75	41
	Non-symptomatic	3.77	5.00	13
	Symptomatic	9.96	3.11	28
Severe Depression	Normative	0.43	0.93	166
	Hostage	2.93	3.39	41
	Non-symptomatic	0.92	2.50	13
	Symptomatic	3.86	3.37	28
Full Scale	Normative	8.47	3.59	158
	Hostage	26.22	14.05	41
	Non-symptomatic	12.61	13.54	13
	Symptomatic	32.54	8.97	28

Discussion

The preliminary analyses supported the argument that care should be taken to re-establish the validity of a measure used within a cultural context different from that in which originally it was developed. They also showed the need for caution when an English language version of a questionnaire is used even with respondents that have English as a second language.

While cross-culturally the responses to the GHQ(20) overall indicated a high degree of consistency, the two items #2 and #18 raised problems. Subsequent attempts to clarify these items statistically were not successful, with loadings of 0.01 and 0.18 respectively obtained on the first (unrotated) principal component, and further attempts to clarify their derived meaning on interview with a representative sample of participants were not fruitful. Therefore it was decided to regard these items as fillers, rather than to discard them, because with other samples of participants in other countries the GHQ20 item measure had proved useful.

structure of the GHQ(20). The difference between the observed data and the four factor model gave a marginally acceptable X^2 value of 22.99, $p < .10$, while at 1.64, the $X^2:df$ ratio was well below the generally accepted upper level of 2.00. Similarly the GFI at 0.97 and the RMSEA at 0.06 both came within the acceptable range of values in confirming the model.

The observed data showed even stronger indications of goodness of fit to the single factor model. There the X^2 value of 0.28 indicated that the data were clearly not significantly different from the model ($p < .90$), and the $X^2:df$ ratio of 0.14 was well below the criterion level of 2.00. These indices were in turn supported by both the GFI that achieved its maximum of 1.00, and the RMSEA that achieved its minimum of 0.00.

Consequently the results were taken as confirmation of both the subscale and full-scale factor structures.

Subscale and Full Scale Characteristics

The relevant scales gave acceptable alpha and split-half coefficients, ranging from .60 to .85, and when set

out in percentile ranks showed an acceptable distribution of total scores.

Clinical Validation

Comparison of the mean scores of the Normative and Hostage groups showed significantly higher level of symptoms in response to the hostage crisis. On all four subscales and on the full scale, the difference between these groups was significant beyond the .001 level.

Comparisons between the mean scores of the normative group with those of the non-symptomatic hostage group and the symptomatic hostage group on each of the subscales and on the full scale, gave considerable support to the validity of the measure as a clinical screening device. The clinical division into symptomatic and non symptomatic groups was reflected in significant differences between the groups on all five measures, beyond the .001 level on four measures, and beyond the .01 level on the fifth (Severe Depression). At the same time, the mean scores of the Non Symptomatic group, although indicating that these participants were significantly less stressed than the Symptomatic group, also returned higher levels of stress than the normative group on all five measures, with Sleeping Problems ($p < .05$) and Anxiety ($p < .01$) significantly so.

Conclusion

The outcome confirmed the robust psychometric properties of the GHQ(20), and validated its use as a supplement to the standard clinical interviews that clinicians and researchers employed within the target population. It also supported the future use of the measure by health professionals in a country in which such aids to assessment have previously been lacking.

References

1. Arbuckle J.L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Small Waters Corporation. Chicago IL.
2. De Girolamo, G., & McFarlane, A.C. (1996). *The epidemiology of PTSD: A comprehensive review of international literature*. In A.J. Marsella, M.J.

Friedman, E.T. Gerrity, R.J. Scurfield. (Eds.). *Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical implications*. Washington DC: American Psychological Association: 33.

3. *Diagnostic and statistical manual IV TR*. (2000). American Psychiatric Association. Washington DC.
4. Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire*. Maudsley Mon. 21 Oxford University Press.
5. Goldberg, D. (1978). *Manual of the General Health Questionnaire*. NFER, Windsor.
6. Goldberg, D., Williams, P. (1988). *A user's guide to the General Health Questionnaire*. NFER-NELSON Windsor.
7. Kishton, J.M., & Widaman, K.F. (1994). Unidimensional versus domain representative parcelling of questionnaire items: an empirical example. *Education & Psychological Measurement*, 54, 757-765
8. Sharma, N. (2001). Politics and primary care: *A focus on Fiji*. Fiji GP, 8, 1, 676-677.
9. Siegert, R.J., McCormick, I.A., Taylor, A.J.W., & Walkey, F.H. (1987). An examination of the reported factor structures of the General Health Questionnaire and the identification of a stable replicable structure. *Australian Journal of Psychology*, 39 1:89-100.
10. SPSS for Windows, Rel. 10.0.0. (1999) SPSS Inc. Chicago IL
11. Taylor, A.J.W., Nailatikau, E., & Walkey, F.H. (2002). A hostage assignment in Fiji. *Australasian Journal of Disaster & Trauma Studies* <http://www.massey.ac.nz/~trauma/issues/2002-2/taylor.htm>
12. Young MA (2001) *The community crisis response team training manual*. (2nd edn). Office of Department of Justice. Washington DC.

The outcome confirmed the robust psychometric properties of the GHQ(20), and validated its use as a supplement to the standard clinical interviews that clinicians and researchers employed within the target population.