

A psychometric evaluation and cross-cultural validation of the 21 item Hopkins Symptom Checklist

Abstract: Data relating to the standardisation of a short version of the 21 item Hopkins Symptom Checklist (HSCL 21) is presented. It confirms the presence of three factors, and provides group means and percentile scales that enable comparisons to be made between the scores of a non-clinical sample and those of two separate target groups that had been under stress. These comparisons give confidence that the rating scale is appropriate for the purpose for which it was used, and suggest that the HSCL 21 could be a useful tool for clinicians in Fiji to use as an adjunct to standard interviews in the assessment of the effects of trauma.

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

Introduction

Long ago Meehl and Rosen (1955) urged clinicians to collect local prevalence based-rate data on indigenous populations systematically, and to validate any instruments they might use in the process of making diagnoses. Choca, Stanley, and Van Denburg (1995, p.30) repeated the plea more recently, with specific regard to the assessment of posttraumatic stress disorder. Then Keane, Kaloupek, and Weathers (1996) asserted that 'there are no assessment instruments that can be used across all cultures and languages...(and) there is much work to be done and much new knowledge needed before we can conduct successfully true cross-cultural studies'. Most recently Kline (2000 p. 361) made the point that '(at) present, unequivocally valid psychometric comparisons across cultures are not possible, although if used with caution, tests can yield some interesting results.'

With such reservations in mind, the present authors present the third in a series of psychometric papers to complete the cross-cultural validation of rating scales that were used on a clinical project in Fiji (Taylor, Aghanwa, & Walkey, unpublished; Aghanwa, Walkey, & Taylor, un-

published). Although the exigencies of the moment precluded the standardisation of the rating scales in advance, local clinicians thought it appropriate for them to be used tentatively in association with clinical interviews when making conventional diagnostic assessments of some hostages and their families that might have been suffering traumatic stress as a result of the May 2000 coup in Fiji (cf. Taylor, Nailatikau, & Walkey, 2002).

In presenting the outcome now, the reliability and the factorial construction of the clinical questionnaires were kept in focus, while taking care not to 'make obeisance to psychometric ritual' (as Kline, 2000 p.333 so elegantly described the inappropriate extent to which some researchers are impelled to go for the sake of convention), and to exemplify his dictum that the actual items in rating scales are often of greater interest than the variables to which they might contribute.

Although the exigencies of the moment precluded the standardisation of the rating scales... , local clinicians thought it appropriate for them to be used ... when making conventional diagnostic assessments of some hostages and their families that might have been suffering traumatic stress as a result of the May 2000 coup in Fiji

The measure

The Hopkins 21 item Symptom Checklist (HSCL 21) was used in the study, together with the General Health Questionnaire (GHQ 20) (Goldberg, 1972; 1978) and the Impact of Event Scale (IES) (Horowitz, Wilner, Alvarez, 1979). It was

thought that the brace of self-report scales together with clinical interviews would give more reliability to diagnoses than might any of the single procedures alone. Furthermore, as Bonnie L. Green (1991) commented, the measures nominated are 'widely used, and provide excellent data for comparing one study with the next.'

The particular short form of the HSCL under consideration was developed by Dianne E. Green and colleagues (Green, Walkey, McCormick, & Taylor, 1988) from the original 58 item version prepared by Derogatis, Lipman, Rickels, Uhlenhuth, & Civi (1974). It was easy to administer and score, had items that loaded on the three factors of general feelings of distress, somatic distress, and performance difficulty (unlike longer versions - see Cyr, McKenna-Foley, & Peacock, 1985), and had been used cross-

*School of Psychology, Victoria University, PO Box 600, Wellington, New Zealand. ** Fiji School of Medicine, Suva, Fiji Correspondent: Dr.F.H.Walkey

culturally (Chung, Bemak, & Wong, 2000). Other forms of the checklist available – viz. the 25item War Zone version (Weathers, Litz, Keane, et.al. 1996) and the variants reported by Norris and Riad (1997) – but none seemed to be superior to the one with which we were acquainted.

The procedure

By way of standardization, the HSCL 21 was administered by the second author to a normative sample of 197 members of hospital staff and medical students in Fiji. None of the respondents was a clinical patient or regarded as being in need of psychological treatment, nor were any directly related to the groups involved in the earlier clinical study. However, like all members of the wider Fijian community at the time, they were unlikely to have completely escaped the effects of 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).

Then confirmatory analyses were made to examine the adequacy of the three-factor structure and the single general factor. The purpose of each analysis was to seek confirmation that the actual response data observed was not unacceptably different from that which could be expected from a theoretical model. Each confirmatory analyses employed four of the highly correlated indicators of the 'goodness of fit' commonly employed to examine the relationship between the response data and the model under scrutiny. In the process it should be noted that:

1. The Chi Square was used to evaluate the significance of differences between the proposed model and the observed data. But contrary to the manner in which the results of statistical analyses are normally interpreted, in the present instance, to be acceptable the differences (which are differences between the observed and the theoretical models) must be small, and therefore not statistically significant.
2. Because many sets of appropriately correlated single item data produce a significant chi square value, the

convention was followed that ideally, the ratio of the Chi Square value to the degrees of freedom (df) should be less than 2.00. There is no specific statistical test to determine the significance either of this ratio, or of 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.

To minimise the effects of error variance in these analyses, two factor-based parcels of items for each factor were included in the examination of the three-factor model, and four domain parcels, each of which contained items from across all three factors, were examined in the one factor model (Kishton & Widaman, 1994).

The confirmatory factor analyses were followed by calculations of the means, standard deviations, and two estimates of reliability (Cronbach's Alpha and corrected split-half correlations)

for each subscale and for the full scale. Finally, a table of percentile norms was created from the responses of this community sample.

The validity of the Scale for use in Fiji was assessed by comparing the HSCL 21 scores of the normative group with those of two groups involved in the earlier hostage crisis that had been identified as either symptomatic or non-symptomatic.

Results

1. Exploratory factor analysis

The initial unrotated principal component loadings ranged from .08 to 0.68, raised doubts about the wisdom of using the sum of responses to all the items as an indicator of scores on a general factor underlying the measure. However because the three items that loaded at low levels on this component appeared also to load significantly on the same rotated and unrotated (second) components, it was decided that these items should be retained in the measure. Consequently all 21 items were included in subsequent analyses. But further assessment of the general factor was left to the subsequent Confirmatory Factor Analyses.

The validity of the Scale for use in Fiji was assessed by comparing the HSCL 21 scores of the normative group with those of two groups involved in the earlier hostage crisis that had been identified as either symptomatic or non-symptomatic.

Table 1. Norms, Summary Statistics and Reliabilities of the Two Sub-scales and of the full Fifteen Item Impact of Events Scale

Percentile Subscale	Intrusion Subscale	Avoidance Scale	Full
1 - 10	0 - 1	0	0 - 2
11 - 20	2 - 3	1 - 4	3 - 10
21 - 30	4 - 5	5 - 7	11 - 14
31 - 40	6 - 7	8 - 9	15 - 17
41 - 50	8 - 9	10	18 - 20
51 - 60	10	11 - 12	21 - 22
61 - 70	11	13	23 - 24
71 - 80	12 - 13	14	25 - 27
81 - 90	14	15 - 18	28 - 31
91 - 95	15 - 16	19 - 20	32 - 34
96 - 100	17 - 35	21 - 40	35 - 75
<i>Mean</i>	<i>8.72</i>	<i>10.32</i>	<i>19.08</i>
<i>Standard Deviation</i>	<i>5.09</i>	<i>5.97</i>	<i>10.04</i>
<i>Alpha</i>	<i>0.84</i>	<i>0.84</i>	<i>0.90</i>
<i>Split Half</i>	<i>0.76</i>	<i>0.80</i>	<i>0.83</i>

2. Confirmatory factor analyses

Evaluation of the adequacy of the three-factor structure, using six-factor based groups of items, showed it to be 'reasonably adequate'. The Chi square value was 12.06, $df=6$, $p<.06$; while the Chi square to df ratio was also 2.01 and only marginally above the rule of thumb maximum of 2.00. The GFI was 0.98, which may be regarded as quite adequate, while the RMSEA was 0.07, which meets the generally accepted criterion for this particular index.

The evaluation of the general factor underlying the measure, utilising seven groups of items, each drawn from all three factor groups, yielded less positive results. The Chi square value was 70.41, $df=14$, $p<.001$. The Chi square to df ratio was considerably in excess of the 2.00 criterion at 5.03, and both the GFI at 0.92 and the RMSEA at 0.13 were also well beyond the criterion for acceptance.

3. Psychometric Characteristics

A summary of psychometric characteristics for the three subscales and for the full scale is given in Table 1.

4. The validation study

Results of the validation study, which enables comparisons to be made between the Normative, the Non-Symptomatic and the Symptomatic groups, are summarised in Table 2.

The results show that differences between the mean scores of the Symptomatic and Non-symptomatic groups on the three subscales and on the Full HSCL 21, were all significant and in the expected direction, as were the comparisons between the results of the symptomatic and normative groups. Comparison of the Non-symptomatic and Normative groups also shows the Non-symptomatic group scoring significantly lower in all four cases. In fact

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

Scale	Group	Mean	Standard Deviation	N
Intrusion	Normative	8.72	5.09	197
	Non-symptomatic	9.62	7.92	13
	Symptomatic	26.43	7.29	28
Avoidance	Normative	10.32	5.97	197
	Non-symptomatic	8.38	10.92	13
	Symptomatic	23.21	8.46	28
Full Scale	Normative	19.08	10.04	197
	Non-symptomatic	18.00	15.36	13
	Symptomatic	49.64	12.44	28

the mean response of the Symptomatic group to the HSCL 21 lay within the range of the 60th to the 80th percentile.

Discussion

The initial exploratory analysis of the HSCL 21 confirmed the adequacy of the item pool as a whole, and the confirmatory procedures supported the three-factor and three-subscale structure of the measure. Supporting the initial observation that three of the items did not load significantly on the first unrotated principal component, the confirmatory analyses provided little support for the use of the full scale. However, given the initial observation, and the relatively high level of reliability shown, it appeared reasonable to retain the possibility of using the full scale, albeit tentatively at the present time. Consequently, norms and percentiles were derived and presented for the full scale as well as for the three subscales.

Though generally adequate, it may be noted that estimates of the reliabilities of the subscales varied considerably, from 0.62 (Corrected Split Half reliability of the Performance Difficulty scale up to 0.84 (Alpha reliability for the Somatic Distress scale). Given the small number of seven items in each scale, the reliability is satisfactory. The reliability estimates for the full scale (0.85 and 0.84) were also satisfactory.

The validation study involved comparisons of the responses of two groups differentially classified following clinical interviews at the time of the earlier hostage crisis in Fiji as either symptomatic (showing evidence of stress) or non-symptomatic. The prediction was, that the symptomatic group would score highly on the HSCL 21 while the Non-symptomatic group would score significantly lower. Scores of these groups are summarised in Table 2. In fact, as predicted, the symptomatic group scored significantly higher than both the non-symptomatic group, and the normative group. Clinicians could therefore have confidence in the measure.

Confidence in the validity of the measure was further supported by the comparison of the scores of the non-symptomatic and the normative groups. As expected, the non-symptomatic group, which had been selected after clinical evaluation from a larger group involved either as hostages or families of hostages, scored significantly lower than the normative group on all three subscales and (consequently) on the full scale. From this it can be said that the scores on the HSCL 21 appear therefore to concur to a very satisfactory degree with the outcomes of clinical evaluation, and the measure may be expected to prove to be a useful screening instrument within the Fiji cultural context.

Conclusions

The three-factor HSCL 21 has proved to be a satisfactory instrument, providing valid and at least adequately reliable indications of symptoms in the three discriminable areas of Performance difficulty, General feelings of Distress and Somatic Distress within the Cultural context of Fiji. But the use of the full scale score, on the other hand, is to be regarded as at least marginally problematic.

References

- Aghanwa, H., Walkey, F.H., & Taylor, A.J.W. (unpublished). The psychometric cross-cultural validation of the Impact of Event Scale
- Arbuckle J.L., & Wothke, W. (1999). *Amos 4.0 User's Guide*. Chicago IL: Small Waters Corporation.
- Chi-Ying Chung, R., Bemak, F., & Wong, S. (2000). Vietnamese levels of distress, Social support, and acculturation: Implications for mental health counseling. *Journal of Mental Health Counseling*, 22, 2, 150-
- Choca, V.P., Stanley, L.A., & Van Denburg, ER. (1995). *Interpretative guide to the Millon Multilevel Inventory*. Washington DC: American Psychological Association.
- Cyr, J.J., McKenna-Foley, J.M., & Peacock, E. (1985). Factor structure of the SCL-90-R: Is there one? *Journal of Personality Assessment*, 29, 6, 571-577.
- 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.
- Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., & Civi, L. (1974). The Hopkins Symptom Checklist(HSCL): A self-report symptom inventory. *Behavioural Science*, 19, 11-14.
- Diagnostic and statistical manual IV.TR.* (2000). Washington DC: American Psychiatric Association.
- Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire*. Maudsley Monograph 21. Oxford University Press.
- Goldberg, D. (1978). *Manual of the General Health Questionnaire*. NFER, Windsor.
- Green, D.E., Walkey, F.H., McCormick, I.A., & Taylor, A.J.W. (1988). Development and evaluation of a 21-item version of the Hopkins symptom checklist with New Zealand and United States respondents. *Australian Journal of Psychology*, 40: 61-70.
- Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of event scale: A measure of subjective stress. *Psychosomatic Medicine*, 41, 209-218.

- Keane, T.M., Kaloupek, D.G., & Weathers, F.W. (1996). Ethnocultural considerations in the assessment of PTSD. In A.J. Marsalla, M.J. Friedman, E.T. Gerrity, & R.M. Scurfield. (1996). *Ethnocultural aspects of posttraumatic stress disorder*. (pp.183-205). Washington DC: American Psychological Association.
- Kishton, J.M., & Widaman, K.F. (1944). Unidimensional versus domain representative parcelling of questionnaire items: an empirical example. *Education & Psychological Measurement*, 54, 757-765
- Kline, P. (2000). *Handbook of psychological testing*. (2nd edn.). London: Routledge.
- Marsalla, A.J., Friedman, M.J., Gerrity, E.T., & Scurfield, R.M. (1996). *Ethnocultural aspects of posttraumatic stress disorder*. Washington DC: American Psychological Association.
- Meehl, P.E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194-216.
- Norris, F.H., & Riad, J.K. (1997). Standardized self-report measures of civilian trauma and post-traumatic stress disorder. In J.P. Wilson & T.M. Keane. (eds.). *Assessing psychological trauma and PTSD*. (pp. 43-68). New York: Guilford Press.
- Sharma, N. (2001). Politics and primary care: A focus on Fiji. *Fiji GP*, 8, 1, 676-677.
- Taylor, A.J.W., Aghanwa, H., & Walkey, F.H. (unpublished). Fiji norms for the General Health Questionnaire.
- 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>
- Weathers, F.W., Litz, B.T., Keane, T.M., Herman, D.S., Steinberg, H.R., Huska, J.A., & Kraemar, H.C. (1996). The utility of the SCL-90-R for the diagnosis of war-zone related posttraumatic stress disorder. *Journal of Traumatic Stress*, 9, 1, 111-128. ■

I've got a great ambition to die of exhaustion
rather than boredom.

Angus Grossant