The use of the General Health Questionnaire as an indicator of mental health in occupational studies

Chris W Clegg

The use of the General Health Questionnaire as an indicator of mental health in occupational studies

MICHAEL H. BANKS,* CHRIS W. CLEGG, PAUL R. JACKSON, NIGEL J. KEMP, ELIZABETH M. STAFFORD AND TOBY D. WALL

MRC Social and Applied Psychology Unit, University of Sheffield

This paper examines the psychometric properties of an existing measure of mental health, the GHQ-12, as revealed in three studies involving employees in an engineering firm ($n = 659$), recent school-leavers ($n = 647$), and unemployed men ($n = 92$). The measure was shown to be psychometrically sound in all cases, with a Likert scoring method providing a more acceptable distribution of scores than the more commonly deployed 'GHQ score' for use in parametric statistical analyses. Scores on GHQ-12 were found to be much higher (indicating lower mental health) for those who were unemployed, higher for women than for men in one sample, and unrelated to age, job level and marital status.

The term mental health has an interesting history within the general area of occupational psychology, having played a more important role at the theoretical level than in empirical studies. The term, for example, features prominently in the writings of Argyris (1951, 1964), Herzberg (1968), and Warr & Wall (1975). Nevertheless it has not given rise to a distinct and coherent set of empirical studies *directly* concerned to identify its antecedents and determinants.

This assertion requires qualification, since its accuracy depends upon how one views the concept of mental health. In practice two main uses of the term are evident in the occupational literature (cf. Murrell, 1978). The first is the notion of '*positive mental health* '(Argyris, 1951; Allport, 1958; Jahoda, 1958), which at an empirical level found its best-known expression in the work of Kornhauser (1965). This refers to behaviours, attitudes and feelings that represent an individual's level of personal effectiveness, success and satisfaction. It has no necessary connection with mental illness in a clinical sense. However, because of the rather vague and general nature of positive mental health thus defined, other authors concerned with this concept have drawn upon a wide range of empirical studies in order to examine its relevance as a dependent variable in occupational settings. In reviews by Cooper & Marshall (1976) and Kasl (1973), for example, one finds consideration given to studies which other authors would categorize as pertinent to job satisfaction, role ambiguity and conflict, trust, stress, and so on. In short, positive mental health has typically been treated at an empirical level through the examination of surrogate measures, which the very generality of the concept encourages. As a result, the usefulness of the concept is called into question, since it could be regarded at best as a higher-order construct relating to one or more traditional areas of inquiry, or at worst an unnecessary semantic elaboration which serves to confound rather than elucidate important issues.

The second use of the term mental health is associated more directly with clinical or medical usage, being defined in terms of the *absence of mental illness*. It is particularly with

*Requests for reprints should be addressed to Dr M. H. Banks, MRC Social and Applied Psychology Unit, University of Sheffield, Sheffield S10 2TN.

respect to this meaning of mental health that occupational psychologists have eschewed empirical research with but a few exceptions. For example, Fraser (1947) examined the incidence of neurosis in a large sample of factory workers and identified job characteristics as among the more important correlates; Erikson *et al.* (1973) found that status congruence was strongly related to the incidence of psychiatric disorder, as indexed by in-patients, out-patients and controls among Navy ratings; and Arthur & Gunderson (1965) found that promotional lag was related to psychiatric illness. Nevertheless, such studies are few and far between.

It may be that this lack of empirical work results from the difficulty Smith (1961) identifies when observing that ' "mental health", and its complement "mental illness" are terms that embarrass psychologists . . . Unable to define or conceptualize them to our satisfaction, we use the term in spite of ourselves '. Also of relevance, however, with respect to mental health in its second sense (the absence of mental illness), is a lack of short demonstrably valid measures of known psychometric acceptability for use with work populations. Whilst such measures exist for job satisfaction, felt stress, role conflict, and so on, it is more difficult to locate equivalent scales of mental health. Yet measures of mental health developed outside the area of occupational psychology exist, and are potentially applicable (see, for example, Eden, 1975). One such measure is the General Health Questionnaire (GHQ) (Goldberg, 1972); what is needed is demonstration of its psycho-metric adequacy within occupational settings together with supporting normative data to enhance its usefulness. The purpose of this paper is to report on the properties of this particular instrument as revealed in three recent empirical studies conducted in work-related settings. First, however, it is necessary to provide some background information on the GHQ itself.

The General Health Questionnaire

The GHQ was designed as a self-administered screening test for detecting minor psychiatric disorders among respondents in community settings. The original development of the measure (Goldberg, 1972) resulted in a 60-item version (GHQ–60) with the ' best ' 30, 20 and 12 of these items (GHQ–30, GHQ–20 and GHQ–12, respectively) being identi-fied for use where respondents ' time is at a premium. Each item retained for scale use was one which showed a difference in endorsements of at least 40 per cent between the propor-tion of ' normals ' and ' severes ' defined by clinical assessments. Items (see instances in Table 1) consist of a question asking whether the respondent has recently experienced a particular symptom or item of behaviour rated on a four-point scale. (Detailed descriptions of the wording of each scale point are given in Goldberg, 1972). Two main scoring methods are used: the ' GHQ–method ' where individuals score 0 if endorsing either of the first two categories, or 1 for endorsing either the third or fourth category; and the ' Likert-method ' where responses are given weights of 0, 1, 2 and 3. The development studies showed that the full scale exhibited high internal consistency and good retest reliability over a period of 6 months. All versions of the scale also correlated highly with one another. The original and subsequent empirical studies (summarized in Goldberg, 1978) have provided evidence of the validity of the GHQ as shown by its linear associations with independent clinical assessments (typically $r = 0.70$ or greater), and its ' sensitivity ' and ' specificity ' in discriminating between ' cases ' and ' normals '. The GHQ-method of scoring was only slightly less satisfactory than the Likert-method in these latter respects.

Examinations of the factor structure of the GHQ have typically yielded a large general factor, with three more subsidiary ones, the general factor being equally evident with the larger (GHQ–60) and shorter (GHQ–12) version of the measure. A 28-item version (GHQ–28) based on a four-factor solution of GHQ–60 has been offered for use (Goldberg, 1978) and is recommended for studies in which more information is required than a single severity score. Nevertheless, because of the underlying general factor a single severity score

Table 1. *Corrected item-whole correlations and reliability coefficients for three samples by each scoring method*

	Sample A[a] (n = 633: one company)		Sample B (n = 512: school-leavers)		Sample C (n = 92: unemployed)	
	GHQ	Likert	GHQ	Likert	GHQ	Likert
Have you recently						
(1) been able to concentrate on whatever you're doing?	0·53	0·45	0·52	0·47	0·53	0·59
(2) lost much sleep over worry?	0·38	0·50	0·38	0·42	0·50	0·52
(3) felt that you are playing a useful part in things?	0·28	0·26	0·44	0·43	0·38	0·51
(4) felt capable of making decisions about things?	0·43	0·31	0·27	0·27	0·55	0·60
(5) felt constantly under strain?	0·44	0·47	0·51	0·52	0·66	0·72
(6) felt you couldn't overcome your difficulties?	0·46	0·51	0·39	0·43	0·64	0·73
(7) been able to enjoy your normal day-to-day activities?	0·58	0·52	0·44	0·46	0·59	0·64
(8) been able to face up to your problems?	0·50	0·40	0·45	0·38	0·68	0·75
(9) been feeling unhappy and depressed?	0·68	0·68	0·57	0·60	0·56	0·67
(10) been losing confidence in yourself?	0·47	0·56	0·57	0·62	0·59	0·67
(11) been thinking of yourself as a worthless person?	0·40	0·45	0·50	0·53	0·31	0·35
(12) been feeling reasonably happy all things considered?	0·59	0·53	0·55	0·52	0·60	0·68
Mean corrected item-whole correlations:	0·48	0·48	0·47	0·48	0·56	0·63
α-coefficients	0·90	0·82	0·82	0·82	0·87	0·90

[a] Analyses are based only on the sample for whom complete data were available.

is derived from the GHQ–28, and this index is equivalent to the GHQ–60 and the GHQ–30 with respect to its performance against independent clinical assessments.

While the evidence supporting the GHQ is both consistent and persuasive it is not in a form fully appropriate for use in occupational studies. In particular, little direct evidence has been recorded for the GHQ–12, which is the version most relevant in occupational studies where mental health will typically be measured as one of several variables and where the use of longer versions of the GHQ is precluded by time considerations. It needs to be determined whether this short version of the GHQ shows adequate psychometric properties with respondents in employment studies, as a first step in the accumulation of normative material which is essential to diagnostic and comparative research. Moreover, the choice of scoring methods requires examination. Whereas the GHQ-method, which is essentially binary, has proved adequate with respect to discriminating between ' cases ' and ' normals ', the equally sensitive Likert-method might be preferable in that it is likely to produce a wider and less skewed distribution of scores more appropriate for correlational analyses and intergroup comparisons based on parametric statistics. This paper is directed at providing evidence related to these issues as revealed by three separate studies. Findings relating GHQ–12 to selected demographic and occupational variables are also reported.

METHOD

Data relevant to an examination of the psychometric properties of the GHQ–12 were obtained in the course of three separate investigations. The samples upon which these are based are described below.

Sample A consisted of employees in an engineering plant. The total population of some 2800 was stratified by job level and the sample selected from within each level at random. Anonymously completed questionnaires were obtained from 659 respondents, representing a 75 per cent response rate which was equivalent across all job levels.

Sample B consisted of 16 year old recent school-leavers representative of the lesser qualified individuals from 12 schools in an urban area (Stafford *et al.,* 1979). The 647 completed interviews during which the GHQ was administered resulted from an 80 per cent contact rate, with 95 per cent of those traced agreeing to take part in the research. The analyses to be reported are based upon those in employment, or unemployed but looking for work ($n = 512$). Because of their small numbers, those who had decided to continue their education or training were excluded.

Sample C consisted of 92 unemployed men, selected at random from a city unemployment register. Further details of this study are given in Hepworth (1980).

RESULTS

The findings from these three samples are described in two parts. The first part focuses on the psychometric properties of the scale and the differences between the alternative scoring methods. The second presents the results reflecting on differences in GHQ–12 scores in relation to demographic and occupational variables, in particular employment status, marital status, age, job level and sex.

Psychometric properties of the GHQ–12

Corrected item–whole correlations are presented for both scoring methods and for the three samples in Table 1. These values remained high across the three samples, and there were no differences between the two scoring methods except in the unemployed sample (C) where the Likert-method produced consistently higher item–whole correlations than the GHQ-method. The alpha coefficient, an index of internal consistency (Cronbach, 1951), also shown in Table 1, was acceptably high with all six values falling between 0·82 and 0·90.

Table 2. *GHQ–12 scores in three samples by employment status, marital status and sex, for two scoring methods*

	Employed		Unemployed	
	Sample A	Sample B	Sample B	Sample C
Single				
GHQ-method				
Mean	1·21	1·52	3·85	4·54
SD	2·02	2·18	3·30	3·23
n	114	431	81	28
% > 2	21·1	22·3	58·0	57·1
Likert-method				
Mean	8·98	8·67	14·06	15·29
SD	4·08	5·07	6·79	6·85
n	114	431	81	28
Married				
GHQ-method				
Mean	0·97	—	—	4·58
SD	1·87	—	—	3·95
n	516	—	—	52
% > 2	15·5	—	—	57·7
Likert-method				
Mean	8·71	—	—	15·10
SD	3·95	—	—	8·41
n	516	—	—	52
Male				
GHQ-method				
Mean	1·04	1·27	3·78	4·76
SD	1·89	1·90	3·35	3·66
n	552	241	37	91
% > 2	17·0	20·3	56·8	60·4
Likert-method				
Mean	8·80	7·86	13·84	15·61
SD	4·02	4·26	6·61	7·82
n	552	241	37	91
Female				
GHQ-method				
Mean	0·86	1·85	3·91	—
SD	2·20	2·46	3·29	—
n	83	190	44	—
% > 2	13·0	24·9	59·1	—
Likert-method				
Mean	8·53	9·71	14·25	—
SD	3·65	5·66	7·01	—
n	83	190	44	—

It is concluded, therefore, that the internal homogeneity and the scaling properties of the GHQ–12 are sufficiently good to justify the use of a single scale score. Furthermore, principal components analysis showed a first factor accounting for between 34 and 48 per cent of the variance in the three samples. Further factors contributed little and were not easily interpretable. In all three samples use of the Likert-method showed a slightly stronger first major factor than the GHQ-method. This is in keeping with the finding of Goldberg (1972) that a first major factor accounted for 45·6 per cent of the variance in the items, and is strong evidence for thinking of GHQ–12 as a unidimensional measure of minor psychiatric disorder.

When the frequency distributions derived from each scoring method are considered, small but consistent differences are found in each sample. Both scoring methods give distributions which are positively skewed, as would be expected in a measure of this sort, but the degree of skew is consistently less for the Likert scoring method. This method also leads to kurtosis values which are closer to zero; that is, the Likert-method gives a less peaked distribution than the GHQ-method, and one closer to a normal distribution. In terms of skewness and kurtosis the Likert-method is consistently more 'normal' and therefore more appropriate for parametric multivariate analyses.

GHQ–12 and demographic and occupational variables

It was possible to examine the association between GHQ–12 scores (mean, standard deviation, and percentage of cases scoring greater than two – this is the threshold point recommended by Goldberg (1972) for the GHQ–12) and the following demographic and occupational variables—age, sex, marital status, employment status and job level.

Taking job level and age first of all, a comparison between four job levels (blue-collar, white-collar, supervisory and managerial employees) on GHQ–12 scores in samples A and B showed no association between these two variables for either sample. Likewise for age and GHQ–12 in samples A and C (sample B were all 16 or 17 year olds), there was no statistically significant relationship.

The influence of employment status, sex and marital status on GHQ–12 scores is shown in Table 2. Because of the nature of the samples (sample A were all employed, sample B were all single, and sample C were all unemployed males) a complete cross tabulation involving these three variables was not possible. Instead, a number of two-group comparisons were undertaken using t tests to assess differences. In all comparisons, the two scoring methods led to the same conclusions. There were no significant differences between married and single in samples A and C (see Table 2). However, large differences found between the employed and the unemployed within sample B ($t = 8 \cdot 06$, d.f. $= 510$, $P < 0 \cdot 001$ for GHQ-method; $t = 8 \cdot 36$, f.g. $= 510$, $P < 0 \cdot 001$ for Likert-method); and in a comparison between all the employed in sample A and all the unemployed in sample C ($t = 14 \cdot 36$, d.f. $= 724$, $P < 0 \cdot 001$ for GHQ-method; $t = 13 \cdot 15$, d.f. $= 724$, $P < 0 \cdot 001$ for Likert-method). In all cases, the unemployed scored higher on the GHQ–12. Results relating GHQ–12 to employment status and sex are also shown in Table 2. No sex difference was found in sample A nor for the unemployed group in sample B. There was however a significant sex difference for the young employed group in sample B ($t = 2 \cdot 76$, d.f. $= 429$, $P < 0 \cdot 01$ for the Likert-method), with females scoring higher than males.

DISCUSSION AND CONCLUSIONS

The psychometric properties of the GHQ–12 are clearly adequate as shown by its high internal consistency and unidimensional factor structure across three different samples. The alternative scoring methods do not differ in this respect, but the Likert-method is to be preferred to the GHQ-method in studies using parametric multivariate techniques since its distribution more closely approximates the normal. Given existing validational evidence for the items on which GHQ–12 is based, these findings are compatible with the view that this measure is appropriate for use in employment studies as an estimate of the severity of psychiatric illness in groups or individuals.

The GHQ–12 scores were not sensitive to differences in age, job level or marital status. As far as age is concerned, it may therefore be assumed in employment studies that differences in age within a population are not accounting for significant portions of the GHQ variance. The same is true for job level, although it would be desirable to replicate these findings with a more heterogeneous group of occupations. It was shown that the GHQ–12 is sensitive to sex differences and to differences in employment status. The latter

is particularly important given that various studies have described the negative psychological effects of unemployment (for example, Marsden & Duff, 1975). This finding thus offers some construct validity for the scale. Moreover, further analysis of data on sample B indicated that, although the unemployed scored higher on the GHQ, this relationship was moderated by motivation to work, such that the unemployed with a high motivation to work scored higher than those with a lower work motivation (Stafford *et al.*, 1980).

Another important application of the GHQ-12 as an indicator of mental health is in studies of job redesign. It has already been demonstrated that favourable changes in mental health as measured by GHQ-12 accompany changes in specified features of jobs (Wall & Clegg, 1980) and advances in this aspect of evaluation will be greatly aided by valid and reliable mental health measures.

Finally, it is essential to stress that the GHQ is recommended here, not for case identification, although that is one of its primary purposes (Goldberg, 1972), but for comparing levels of psychiatric illness within and between populations. Used in this way the GHQ provides a useful estimate of the severity of psychiatric illness for use in the study of employment-related and occupational problems.

. REFERENCES

ALLPORT, G. W. (1958). Personality: Normal and abnormal. *Sociological Review*, 6, 167–180.
ARGYRIS, C. (1951). *Personality Fundamentals for Administrators*. Yale: Labor Management Center.
ARGYRIS, C. (1964). *Integrating the Individual and the Organization*. New York: Harper.
ARTHUR, R. J. & GUNDERSON, E. K. (1965). Promotion and mental illness in the navy. *Journal of Occupational Medicine*, 7, 452–456.
COOPER, C. L. & MARSHALL, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology*, 49, 11–28.
CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
EDEN, D. (1975). Organizational membership versus self-employment: Another blow to the American dream. *Organizational Behavior and Human Performance*, 13, 79–94.
ERIKSON, J., EDWARDS, D & GUNDERSON, E. K. (1973). Status congruency and mental health. *Psychological Reports*, 33, 395–401.
FRASER, R. (1947). The incidence of neurosis among factory workers. Industrial Health Research Board, Report no. 90. London: HMSO.
GOLDBERG, D. (1972). *The Detection of Psychiatric Illness by Questionnaire*. London: Oxford University Press.
GOLDBERG, D. (1978). *Manual of the General Health Questionnaire*. Windsor: National Foundation for Educational Research.
HEPWORTH, S. J. (1980). Moderating factors of the psychological impact of unemployment. *Journal of Occupational Psychology*, 53, 139–145.
HERZBERG, F. (1968). *Work and the Nature of Man*. London: Staples Press.
JAHODA, M. (1958). *Current Concepts of Mental Health*. New York: Basic Books.
KASL, S. V. (1973). Mental health and work environment: An examination of the evidence. *Journal of Occupational Medicine*, 15, 509–518.
KORNHAUSER, A. (1965). *Mental Health of the Industrial Worker*. New York: Wiley.
MARSDEN, D. & DUFF, E. (1975). *Workless*. Harmondsworth, Middx: Penguin.
MURRELL, H. (1978). Work stress and mental strain. Occasional paper no. 6, Department of Employment Work Research Unit.
SMITH, M. B. (1961). Mental health reconsidered: A special case of the problems of values in psychology. *American Psychologist*, 16, 299–306.
STAFFORD, E. M., BANKS, M. H. & JACKSON, P. R. (1979). Young people starting work: First survey of Leeds 1978 school leavers. Memo 355, MRC Social and Applied Psychology Unit, University of Sheffield.

STAFFORD, E. M., JACKSON, P. R. & BANKS, M. H. (1980). Employment, work involvement and mental health in less qualified young people. *Journal of Occupational Psychology,* **53** (in press).

WALL, T. D. & CLEGG, C. W. (1980). A longitudinal field study of group work redesign *Journal of Occupational Behavior* (in press).

WARR, P. B. & WALL, T. D. (1975). *Work and Well-being.* Harmondsworth, Middx: Penguin.