Published paper:

Smith, AB, Oluboyede, Y, West, R, Hewison, J and House, AO (2013) *The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress.* Quality of Life Research, 22 (1). 145 – 152.

http://dx.doi.org/10.1007/s11136-012-0133-7

Title: **The factor structure of the GHQ–12: The interaction between item phrasing, variance and levels of distress**

Authors: Adam B. Smith[1], Yemi Oluboyede[2], Robert West[2], Jenny Hewison[2], Allan O. House[2]

Affiliations:

1. York Health Economics Consortium, University of York, UK
2. Leeds Institute of Health Sciences, University of Leeds, UK

Corresponding author's email address: adam.smith@york.ac.uk
Corresponding author's telephone and fax number: +44(0)1904 324831/+44 (0) 1904 323628

Total word count: 3055

Abstract

Purpose: The GHQ–12 is a self-report instrument for measuring psychological morbidity. Previous work has suggested several multidimensional models for this instrument, although it has recently been proposed that these may be an artefact resulting from a response bias to negatively phrased items. The aim here was to explore the dimensionality of the GHQ-12.

Methods: Cluster analysis, exploratory factor analysis and confirmatory factor analysis were applied to waves of data from the English Longitudinal Study of Ageing (ELSA wave 1 and 3), in order to evaluate fit and factorial invariance over time of the GHQ-12.

Results: Two categories of respondents were identified: high and low scorers. Item variances were higher across all items for high scorers and higher for negatively-phrased items (for both high and low scorers). The unidimensional model accounting for variance observed with negative phrasing (Hankins, 2008) was identified as having the best model fit across the two time points.

Conclusions: Item phrasing, item variance and levels of respondents distress affect the factor structure observed for the GHQ-12 and may perhaps explain why different factor structures of the instrument have been found in different populations.

Keywords: GHQ-12, factor structure, factorial invariance, cluster analysis, confirmatory factor analysis, community sample, elderly population, ELSA

Abbreviations

GHQ-12:	General Health Questionnaire – 12
ELSA:	English Longitudinal Study of Ageing
BIC:	Bayesian Information Criterion
EFA:	Exploratory Factor Analysis
PCA:	Principal components analysis
CFA:	Confirmatory Factor Analysis
RMSEA:	Root-Mean-Square Error of Approximation
ECVI:	Expected Cross-Validation Index
ANOVA:	Analysis of Variance
ANCOVA:	Analysis of Covariance

Introduction

The General Health Questionnaire (GHQ) was designed for assessing psychological morbidity in both community and non-psychiatric settings [1]. The instrument has undergone a number of modifications over time from the original GHQ comprising 60 items to a more recent 12-item version, the GHQ–12. The GHQ–12 is a well-validated instrument for assessing psychological morbidity [2], yet despite its brevity there remains considerable debate about the factor structure of the instrument, which has important implications for the use of the instrument, particularly in a clinical context. A number of exploratory and confirmatory factor analysis studies have found evidence for two and three factor structures despite the fact that the GHQ–12 was originally intended as a unidimensional instrument.

Factor analysis studies have proposed a number of two-factor models for the GHQ–12. A principal components factor analysis by Politi *et al*. [3] identified a two-factor structure for the GHQ–12 corresponding to a seven-item ‘General Dysphoria’ factor consisting of the anxiety and depression items, and a six-item ‘Social Dysfunction’ factor, consisting of items relating to daily activities and ability to cope, with one item (item 12, “Not feeling happy”) loading weakly onto both factors. A very similar two-factor structure (Anxiety/Depression and Social Dysfunction with seven and five items respectively) with overall happiness loading only on Social Dysfunction and a decision-making item loading onto Anxiety/Depression has also been determined by another study [4].

Other two-factor models proposed include a six-item Anxiety/Depression factor and a five-item Daily Activities and Social Performance factor [5], as well as two factors representing the domains Depression and Social Dysfunction [2].

A number of three-factor models have also been suggested [6–7], including a structure consisting of three factors (‘Social Performance’, ‘Anhedonia’ and ‘Loss of confidence’) with three cross-loading items (e.g. ‘concentrate’, ‘enjoy normal activities’, and ‘feeling reasonably happy’) [8-9]. In addition to this a significant number of population-based studies have provided support for Graetz’s [10] three-factor model comprising Anxiety / Depression, Social Dysfunction and Loss of Confidence [11–15].

It has been argued that separate factor loadings may be based on item valence with positively and negatively worded items loading onto two separate factors. For instance, an item response analysis (Rasch) of the GHQ-30 suggested a two-factor structure based on positive and negatively worded items respectively [16]. Similarly, subsequent analyses on an 8-item GHQ have also demonstrated a two-factor structure reflecting item valency [17–18]. A recent study has suggested that the various models proposed for the GHQ–12 may be in fact be an artefact caused by a response bias to the negative wording of six of the items with these items creating additional variance not produced by positively phrased questions [19]. The results of the latter study indicated that when higher levels of variance for negatively phrased items were accounted for in a confirmatory factor analysis by allowing error terms on the negatively worded items to correlate (within the three-factor model, for instance Graetz's, these "negative" items are distributed across the anxiety/depression and confidence factors), this resulted in better model fit for a unidimensional structure for the GHQ–12, in comparison to other two and three factor structures. Some support has recently been found [20] for this in a study comparing this unidimensional model with Graetz's three factor, as well as a two-factor model (the Andrich and van Schoubroeck model), where the unidimensional model proved to have marginally better fit compared to the other two models.

As is evident from the above there remains uncertainty surrounding the factor structure of the GHQ–12. The high degree of correlation reported between factors has often led a number of authors to recommend using the summed GHQ–12 scores [12–13] despite the various multidimensional models proposed. The factor structure has important implications for the clinical use in interpreting scores and for how the GHQ–12 should be used to identify psychological morbidity. This is particularly the case for an ageing population given the fact that mental health (such as anxiety and depression) is frequently both under-diagnosed and under-treated in the elderly both in the community as well as in residential care [21–22].

Therefore, the aims of this study were twofold: firstly to apply a cluster analysis to the GHQ–12 completed by a large (*N*=6237) representative sample of the ageing population (aged 50+) to ascertain whether there are subgroups of respondents, and secondly to explore the factor structure of the GHQ–12 in particular to evaluate, at two time points, a number of competing models for any potential subgroups identified in the cluster analysis to test the stability of the models.

Method

This study utilises data from Wave 1 and Wave 3 of English Longitudinal Study of Ageing (ELSA). ELSA is a longitudinal survey of individuals aged 50 and over. It covers a diverse range of topics necessary to understand the economic, social, psychological and health elements of the ageing process. The sample in this study comprises of individuals who had completed data at Wave 1 and Wave 3 (N=6237). The first and third waves of ELSA were collected between March 2002 – March 2003 and May 2006 – August 2007 respectively (see Marmot *et al* [23], for further information on Wave 1). The two waves of data from the ELSA were used to compare the stability of the GHQ-12 factor structure models.

The GHQ–12 is a 12-item instrument designed for assessing and detecting psychological morbidity [24]. There are four response categories which can be scored using the original dichotomous scoring system (0-0-1-1), as well as a modified dichotomous system (0-1-1-1). Finally, the GHQ–12 may also be scored as a Likert scale (on a 0-1-2-3 scale). There is evidence to suggest that ordinal, Likert scoring of the GHQ–12 allows better discrimination between competing models in confirmatory factor analyses of the GHQ–12 [25], therefore this scoring method was employed in this study. **However, given that the dichotomous scoring is also commonly used to identify clinical levels of psychological distress this method was also used within the confirmatory factor analysis.**

A two-step cluster (SPSS 15) was applied to the GHQ–12 for Wave 1 and Schwartz Bayesian Information Criterion (BIC) used to determine the number of clusters. Subsequently, variance of the positive and negatively phrased items was determined and an exploratory factor analysis (EFA) was applied to this wave, i.e. the entire sample to establish a benchmark model against which to compare subsequent factor structures. Principal components analysis with varimax rotation was used for the EFA.

A confirmatory factor analysis (CFA) was applied to the Wave 1 sample and the subgroups derived from the cluster analysis using AMOS 7. The CFA compared four models (unidimensional with/out correlated error terms, the two factor structure – positive and negative phrased items, and the three-factor model). Maximum likelihood estimation was used for the CFA. The goodness-of-fit of each model was assessed using the Sattora–Bentler scaled chi-square, the Comparative Fit Index (CFI, [26]) and the Incremental Fit Index (IFI,

[27]). Additionally, the Root-Mean-Square Error of approximation (RMSEA, [28]) was included with 90% confidence intervals. Non-significant chi-squares and values greater than 0.95 are considered as acceptable model fit for the CFI and IFI. RMSEA values below 0.08 are considered to reflect acceptable fit to the model and values smaller than 0.05 as good fit [29]. Finally, a comparison of fit between the various models was also included using the Expected Cross-Validation Index (ECVI). The smallest value for the ECVI was used to indicate the best model fit [13].

The CFA was also applied to the second wave (Wave 3) of data for the subgroups defined in the earlier analysis.

Differences between scores by patient cluster and item valence were investigated using a one-way analysis of variance (ANOVA). The partial eta squared values were also recorded to determine the amount of variance explained by the independent variable (cluster).

Results

Wave 1 – Cluster Analysis

The results of the cluster analysis revealed two groups of : high and low scorers. Low scorers tended to score lower on negatively phrased items in respect of positively phrased items, as well as in comparison with high scorers (Figure 1).

Wave 1 – GHQ–12 scores

Item variances for negatively phrased items were higher than for positively phrased items (Table 1). The analysis of variance revealed (ANOVA) statistically significant differences were found between mean scores for the positive and negatively worded items, as well as GHQ–12 total by group: $F(1, 6235) = 1470.82$, $p < 0.0001$; $F(1, 6235) = 9564.56$, $p < 0.0001$; and $F(1, 6235) = 6991.51$, $p < 0.0001$, respectively. The partial eta squared value for the total GHQ–12 was 0.53; the values for the positive and negatively phrased items were 0.19 and 0.61 respectively, suggesting that patient cluster accounted for substantially more of the variance for the negatively phrased compared to the positively phrased items.

This latter finding confirms the earlier results reported by Hankins [19]. In addition, item variances were higher across all items for high scorers in comparison with the low scorers, and item variances were particularly high for this group for the negatively phrased items, suggesting a three-way interaction between item valence, item variance and level of distress as shown by Figure 2.

This suggests that high scorers (i.e. people with high levels of emotional distress) may be disproportionately affected by negatively phrased items. If, indeed, a response bias lies at the heart of explaining the various factor models proposed for the GHQ–12 the unidimensional model accounting for differences in variance would provide the best fit, in particular for the group of high scorers.

Wave 1 – Exploratory and Confirmatory Factor Analysis

The EFA of Wave 1 suggested a two-factor structure consisting of positive and negatively phrased items. The factor structure accounted for 54.9% of the variance explained.

Table 2 shows the results of the Confirmatory Factor Analysis for Wave 1. For the overall sample the best model fit (χ^2 = 903.66, df = 39, $p < 0.0001$) was demonstrated by the unidimensional model with shared variance (Hankins, 2008) based on all four fit criteria (e.g. RMSEA = 0.06, 90%CI 0.056 – 0.063). This model also produced the best fit (on all criteria) for the low scorers (RMSEA = 0.039, 90%CI 0.035–0.044). Graetz's model [10] also revealed good fit statistics for the low scorers with the RMSEA demonstrating overlapping 95% confidence intervals with Hankins' model for this group. Both Hankins' and Graetz's model had the best fit for the high scorers, in particular the former model although the RMSEA fit statistic fell just outside acceptable fit levels (e.g. RMSEA = 0.087, 90%CI = 0.082–0.093). **Results from the CFA using the dichotomous scoring method revealed very similar model fit statistics (not shown).** Taken together the results demonstrated that Hankins' model provided the most optimum fit for all three groups: overall sample, low and high scorers.

Wave 3 – GHQ–12 scores

The means and standard deviations for GHQ–12 scores at Wave 3 are shown in Table 3. It can be seen once again that although negatively worded items produced higher levels of variance in comparison to positively worded items across groups, item variance was considerably higher for the high scorers. Statistically significant differences were again found between mean scores for the positive and negatively worded items, as well as GHQ–12 total by group on the ANOVA: $F(1, 6235) = 3624.32$, $p < 0.0001$; $F(1, 6235) = 10382.46$, $p < 0.0001$; and $F(1, 6235) = 7006.09$, $p < 0.0001$, respectively. The partial eta squared value for the total GHQ–12 was 0.53. The partial eta squared was 0.18 for positively phrased items compared to 0.63 for the negatively phrased items, i.e. patient cluster accounted for greater degree of the variance for the latter. These results confirm a three-way interaction between item valence, item variance and level of distress (Figure 3).

Wave 3 – Confirmatory Factor Analysis

The results of the confirmatory factor analysis at Wave 3 demonstrated (Table 4) that the unidimensional model with shared variance also provided the best fit over time (e.g. Overall

RMSEA = 0.059, CI 90% 0.056–0.063, High scorers RMSEA = 0.064, CI 90% 0.059–0.070, Low scorers RMSEA = 0.056, CI 90% 0.051–0.060). **These results were replicated for the dichotomously scored data with the exception of Graetz's model (1991) for low scorers where only the saturated model fit the data (results not shown).**

Conclusion

Previous studies have suggested a number of competing models for the GHQ–12. The aim of this study was to apply a cluster analysis to GHQ–12 data to determine whether any subgroups of respondents could be identified. Additionally, we aimed to evaluate a number of models previously proposed for the GHQ–12.

The results of the cluster analysis identified two groups of respondents, high and low scorers, **both with different response patterns to the GHQ-12, but** particularly the low scorers **who** tended to score low on negatively worded items. **In contrast to** item variance for the high scorers was particularly high (at both time points).

The results of the exploratory factor analysis suggested a two-factor model consisting of positively and negatively worded items. The results of the **confirmatory factor analysis** demonstrated that the unidimensional model with shared variance provided the best fit across all groups and over time **(including for dichotomous scores).** Accounting for the additional variance observed in the high scorers group improved fit significantly, suggesting that a response bias to negatively phrased items may contribute to the different factor structures proposed for the GHQ–12 (see for example Hankins [19]).

It had been anticipated that any differences observed between the clusters identified in terms of the item variance would be accounted for by the correlated error terms. This would lead to the hypothesis that best fit would be observed for the correlated error term model, particularly for those clusters with high levels of variance on negatively phrased items. The results of the confirmatory factor analysis demonstrated that this was the case: this model (Hankins, [19]) displayed the best model fit at the two time points. A closer inspection of the item variance for both waves reveals that high scorers have high levels of item variance irrespective of item valence. Similarly, average item variance for the negatively worded items for low scorers is comparable to the average item variance for the positively worded items for the high scorers. In other words there was an interaction between item valency and item variance, also explaining why controlling for item variance (for the negatively worded items) has an effect on model fit, particularly for Wave 3. In fact, the high level of variance for the high scorers may suggest there was an issue with variance stability. In order to assess this we re-ran the analysis using a square-root transformation in order to counteract any potential variance instability. The results of this additional analysis revealed the same results with the unidimensional model with shared variance demonstrating the best fit.

Hankins' [19] has suggested that negatively worded items give rise to a response bias. Our study extends these findings by demonstrating that the response bias is potentially moderated by levels of distress. Studies investigating the interaction between cognitive biases and anxiety have demonstrated that anxious individuals display an information processing bias with preferential attention to threatening and ambiguous stimuli [30–31]. We explored this possible link further by running an analysis of covariance (ANCOVA) on the positive and negatively phrased items from Wave 3 with patient cluster as the independent variable and the GHQ-12 total score from Wave 1 as the covariate. The results showed statistically significant results both by patient cluster (positive items: $F(1, 6234) = 747.17$, $p < 0.0001$; negative items: $F(1, 6234) = 7625.49$, $p < 0.0001$) and the GHQ–12 (positive items: $F(1, 6234) = 298.42$, $p < 0.0001$; negative items: $F(1, 6234) = 781.48$, $p < 0.0001$). The partial eta squared values were higher for the negatively phrased items for both GHQ–12 total scores (0.11 for negative items versus 0.05 for positive items) and patient cluster (0.55 versus 0.11). These results add further weight to cognitive bias as a potential explanation for the interaction between item phrasing, item variance and levels of distress.

Taken together this suggest that levels of distress prevalent in a population may potentially affect the factor structure of the GHQ–12 and may therefore also help explain why different models have been proposed and/or discovered with different populations. **If the results are replicated in future studies then this would have significant implications on the use of GHQ-12 in research. Hankins [19] has previously noted that the presence of a response bias to negatively phrased items could have an impact on the interpretation of GHQ-12 data, for instance, in an epidemiological context, as well as impacting on the reliability of the instrument. Evidence of an interaction between the level of psychological distress and negative phrasing further adds to the potential issues, given that the results of this study also suggest that any factor structure imposed on the data (without prior analysis) may well be sample dependent. Taken together these results and those of Hankins [19] suggest that at the very least a lelve of**

circumspection is warranted when analysing and interpreting GHQ-12 data within research.

There may also be implications for clinical practice. The utility of using brief measures such as the GHQ–12 as screening tools has been questioned [32] given the relatively low positive predictive value of these measures (largely due to the low prevalence of conditions severe enough to warrant intervention). Nevertheless some authors have recommended continuing to use a summary index of the GHQ–12 despite the presence of multiple factors [12]. Yet the results of this and other studies recommending a three-factor structure ("social dysfunction", "anxiety/depression" and "loss of confidence") suggest that the GHQ–12 may only have limited utility as a screening instrument when used to produce a single summary score leading to an inflation in the numbers of patients identified as requiring further investigation. Using individual factors from the GHQ–12 to identify more specific emotional disorders may improve sensitivity and reduce the number of false positives, however this needs to be balanced against the likelihood of a reduction in the positive predictive value as prevalence would decrease when moving from identifying general distress to a more specified state (i.e. anxiety or depression). Continued use of the GHQ–12 in summary form will also restrict its use in research through not only a possible misidentification of respondents, but also a misrepresentation of scores. These issues will therefore require further empirical work.

A potential limitation of this study is that the sample consisted of respondents aged 50 and above, and as such is perhaps less representative than other studies (e.g. Hankins, 2008), nevertheless this is mitigated by the large numbers of respondents and stability of the models over time. Furthermore, to our knowledge no other study has applied this methodology to a representative sample of elderly individuals living in the UK. It is a novel approach to identify subgroups based on GHQ scores and to assess the GHQ factor structure based on the subgroups identified. Understanding the psychological morbidity of this population is of particular importance as the elderly are more likely to suffer from long standing co-morbidities [33].

In conclusion, this study extends previous work on the factor structure of the GHQ–12, particularly studies that have focused on item variance [19], and has refined this analysis further by focusing on subgroups of respondents identified in the cluster analysis, i.e. high and low scorers. The latter finding may have therapeutic implications in terms of, for instance, relapse prevention. The study suggests that in addition to the valency of items levels of distress also need to be taken into account when comparing across clusters of respondents, which in turn may have important research implications.

References

1. Goldberg, D.P., & Hillier, V.F. (1979). A scaled version of the General Health Questionnaire. Psychological Medicine 9, 139-145.

2. Werneke, U., Goldberg, D.P., Yalcin, I., & Üstün, B.T. (2000). The stability of the factor structure of the General Health Questionnaire. Psychological Medicine 30, 823 – 829.

3. Politi, P.L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity & factor structure of the 12-item General Health Questionnaire among young males in Italy. Acta Psychiatrica Scandinavica 90, 432-437.

4. Kiliç, C., Rezaki, M., Rezaki, B., Kaplan, I., Özgen, G. (1997). General Health Questionnaire (GHQ12 & GHQ28): psychometric properties and factor structure of the scales in a Turkish primary care sample. Social Psychiatry and Psychiatric Epidemiology 32, 327 – 331.

5. Schmitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. Acta Psychiatrica Scandinavica 100, 462-468.

6. Martin, A.J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. Psychological Reports 84, 927-935.

7. Picardi, A., Abeni, D., & Pasquini P. (2001). Assessing psychological distress in patients with skin diseases: reliability, validity and factor structure of the GHQ – 12. European Academy of Dermatology and Venereology 15, 410-417.

8. Campbell, A., Walker, J., & Farrell, G. (2003). Confirmatory factor analysis of the GHQ12: Can I see that again? Australian & New Zealand Journal of Psychiatry 37, 475-483.

9. Worsely, A., & Gribbin, C.C. (1977). A factor analytic study of the twelve-item General Health Questionnaire. Australian and New Zealand Journal of Psychiatry 11, 269-272.

10. Graetz, B. (1991). Multidimensional properties of the 12-item General Health Questionnaire. Social Psychiatry & Psychiatric Epidemiology 26, 132-138.

11. Cheung, Y.B. (2002). A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. International Journal of Geriatric Psychiatry 17, 739-744.

12. Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S.C., & Cheung, Y.B. (2004). Does the 12-item General Health Questionnaire contain multiple factors and do we need them? Health and Quality of Life Outcomes 2, 63.

13. 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, 231-236.

14. Mäkikangas, A., Feldt, T., Kinnunen, U., Tolvanen, A., Kinnunen, M.L., & Pulkkinen, L. (2006). The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: Evidence from two community-based samples. Psychological Assessment 18, 444-451.

15. Penninkilampi-Kerola, V., Miettunen, J., & Ebeling, H. (2006). A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. Scandanavian Journal of Psychology 47, 431-440.

16. Andrich, D., & van Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. Psychological Medicine 19, 469-485.

17. Kalliath TJ, O'Driscoll MP, Brough P (2004). A confirmatory factor analysis of the General Health Questionnaire-12. Stress and Health, 20, 11-20.

18. Smith AB, Fallowfield LJ, Stark DP,Velikova G, Jenkins V (2010). A Rasch and confirmatory factor analysis of the General Health Questionnaire (GHQ) – 12. Health and Quality of Life Outcomes 8, 45.

19. Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? Clinical Practice & Epidemiology in Mental Health 4, 10.

20. Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. Personality and Individual Differences 46,197–201.

21. Clark, S.A. (1992). Mental illness among new residents to residential care. International Journal of Geriatric Psychiatry 7, 59–64.

22. Harrison, R., Savla, N. & Kafetz, K. (1990). Dementia, depression and physical disability in a London borough: a survey of elderly people in and out of residential care and implications for future developments. Age and Ageing 19, 97–103.

23. Marmot, M., Banks, J., Blundell, R., Lessof, C. and Nazroo, J. (2003) Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing, The Institute for Fiscal Studies: London.

24. Goldberg, D., & Williams, P. (1988). A User's Guide to the General Health Questionnaire. NFER-Nelson: Windsor.

25. Campbell, A., & Knowles, S. (2007). A confirmatory factor analysis of the GHQ12 using a large Australian sample. European Journal of Psychological Assessment 23, 2-8.

26. Bentler, P.M. (1990). Comparative fit indices in structural models. Psychological Bulletin 107, 238-246.

27. Bollen, K.A. (1989). Structural equations with latent variables. Wiley: New York.

28. Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioural Research 25, 173-180.

29. Jöreskog, K., & Sörbom, D. (1981). LISREL V: Analysis of linear structural relationships by the method of maximum likelihood. National Educational Resources: Chicago.

30. Brosan, L., Hoppitt, L., Shelfer, L., Sillence, A., Mackintosh, B. (2011). Cognitive bias modification for attention and interpretation reduces trait and state anxiety in anxious patients referred to an out-patient service: Results from a pilot study. Journal of Behavior Therapy and Experimental Psychiatry, 42(3), 258-264.

31. Macleod, C., Mathews, A., Tata, P. (1986). Attentional bias in emotional disorders. Journal of Abnormal Psychology, 95(1), 15-20.

32. Gilbody, S., House, A., & Sheldon, T. (2001). Routinely administered questionnaires for depression and anxiety: systematic review. British Medical Journal 332, 406–409.

33. Harwood, R.H., Prince, M.J., Mann, A.H. & Ebrahim, S. (1998). The prevalence of diagnoses, impairments, disabilities and handicaps in a population of elderly people living in a defined geographical area: the Gospel Oak project. Age and Ageing 27,707–714.

Figure 1 Mean GHQ–12 scores by cluster for Wave 1

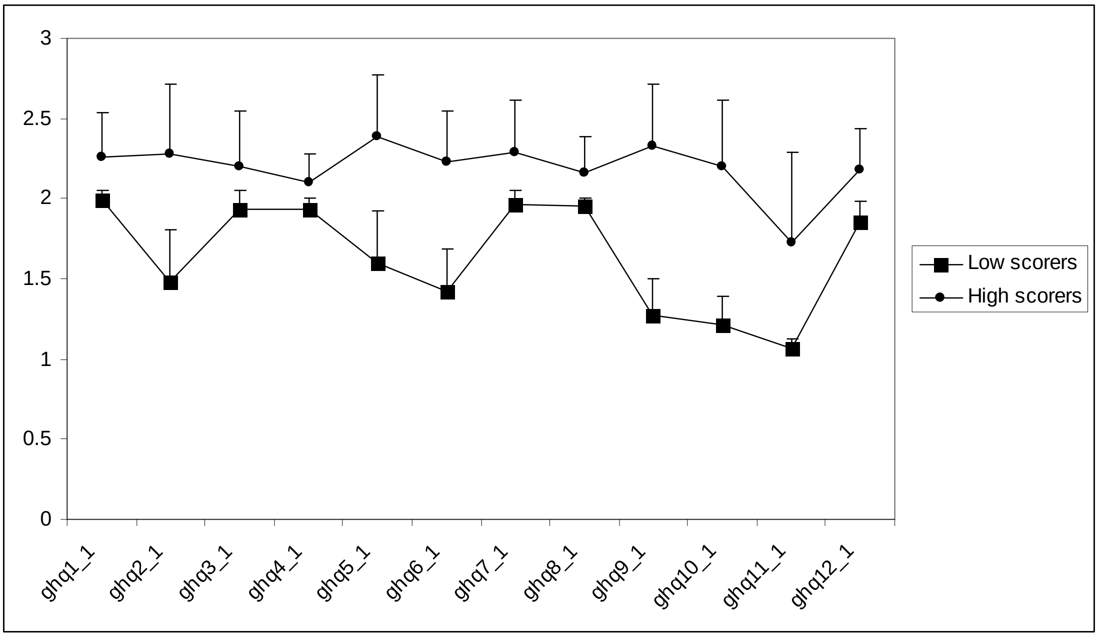

Table 1 Means, standard deviations & variances for GHQ12 (Wave 1)

	Group	N	Mean	Std. Deviation	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
Positive items	low	3830	11.63	1.06	11.60	11.66	6	16
	high	2407	13.20	2.13	13.11	13.28	6	24
Negative items	low	3830	8.05	1.54	8.00	8.10	6	13
	high	2407	13.16	2.59	13.06	13.26	9	24
GHQ total	low	3830	19.68	1.86	19.62	19.74	12	25
	high	2407	26.36	4.35	26.18	26.53	20	48

Figure 2. Item variance for positively and negatively worded items by group (Wave 1)

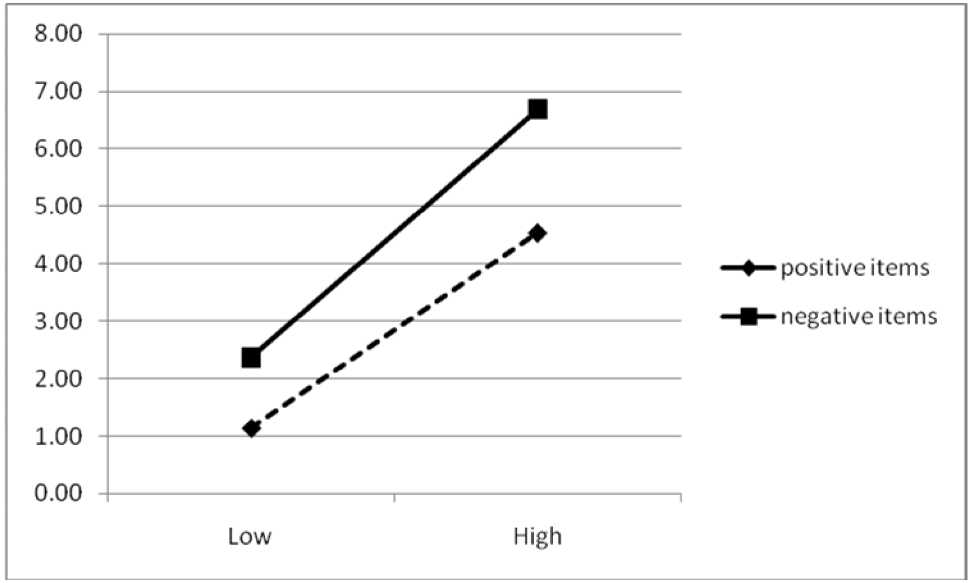

Table 2 – Confirmatory Factor Analysis of the GHQ–12 at Wave 1

Groups	**Overall**				**High**				**Low**			
Models	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
X^2	4651.59	903.656	2327.0	1504.84	1493.13	754.64	1453.29	990.27	1732.30	269.68	615.71	359.15
df	54	39	53	51	54	39	53	51	54	39	53	51
p	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
RMSEA	0.12	0.06	0.08	0.07	0.11	0.087	0.11	0.09	0.09	0.039	0.05	0.04
RMSEA 95%	0.114-0.12	0.056-0.063	0.08-0.086	0.065-0.071	0.101-0.110	0.082-0.093	0.10-0.109	0.083-0.092	0.086-0.094	0.035-0.044	0.049-0.056	0.036-0.044
ECVI	0.76	0.16	0.39	0.25	0.65	0.36	0.64	0.44	0.47	0.097	0.18	0.11
IFI	0.84	0.97	0.92	0.95	0.84	0.92	0.85	0.90	0.60	0.95	0.87	0.93
CFI	0.84	0.97	0.92	0.95	0.84	0.92	0.85	0.90	0.60	0.95	0.87	0.93

Models:
[1]unidimensional
[2]unidimensional with correlated error
[3]two-factor structure (+/-ve items)
[4]three-factor structure (Graetz, 1991)

Table 3 Means, standard deviations & variances for GHQ–12 (Wave 3)

	Group	N	Mean	Std. Deviation	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Positive items	high	2744	13.23	2.19	13.15	13.32	6	24
	low	3493	11.70	1.01	11.66	11.73	6	16
Negative items	high	2744	12.87	2.58	12.77	12.97	7	24
	low	3493	7.68	1.38	7.64	7.73	6	12
GHQ total	high	2744	26.10	4.38	25.94	26.27	20	48
	low	3493	19.38	1.63	19.33	19.43	12	24

Figure 3. Item variance for positively and negatively worded items by group (Wave 3)

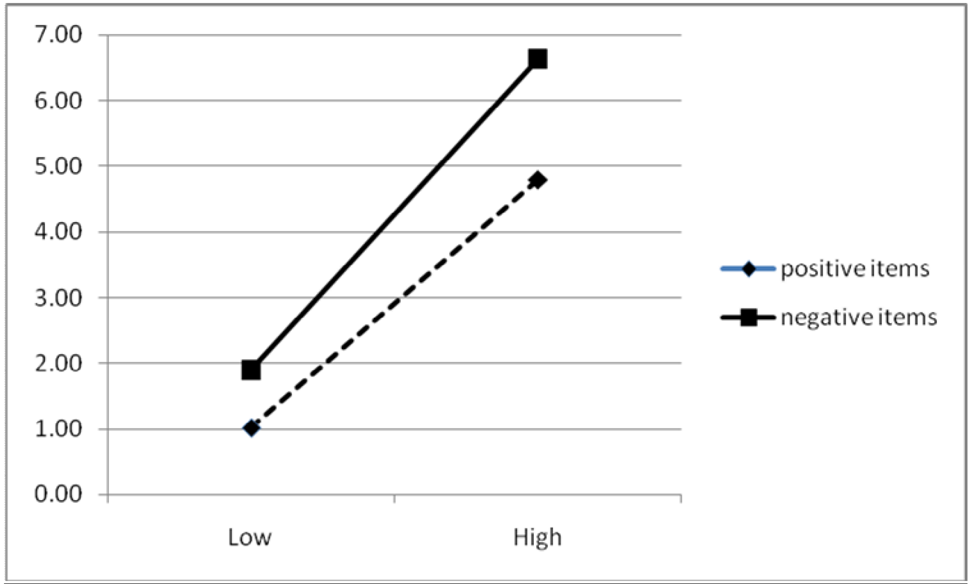

Table 4 – Confirmatory Factor Analysis of the GHQ–12 at Wave 3

Groups	**Overall**				**High**				**Low**			
Models	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
X^2	4414.38	897.46	2249.54	1578.09	1337.00	426.72	930.41	687.58	2831.17	502.59	1318.84	890.46
df	54	39	53	51	54	40	53	51	54	39	53	51
p	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
RMSEA	0.11	0.059	0.082	0.069	0.099	0.064	0.083	0.072	0.116	0.056	0.079	0.066
RMSEA 95%CI	0.11-0.12	0.056-0.063	0.079-0.084	0.066-0.072	0.095-0.104	0.059-0.07	0.078-0.088	0.067-0.077	0.112-0.12	0.051-0.06	0.075-0.083	0.062-0.069
ECVI	0.72	0.16	0.39	0.27	0.59	0.22	0.42	0.44	0.76	0.16	0.36	0.25
IFI	0.85	0.97	0.93	0.95	0.89	0.97	0.93	0.95	0.77	0.96	0.90	0.93
CFI	0.85	0.97	0.93	0.95	0.89	0.97	0.93	0.95	0.77	0.96	0.90	0.93

Models:
[1]unidimensional
[2]unidimensional with correlated error
[3]two-factor structure (+/-ve items)
[4]three-factor structure (Graetz, 1991)