

Manuscript Number: JAD-D-11-00794R1

Title: A hierarchy of distress and invariant item ordering in the General Health Questionnaire-12

Article Type: Short Communication

Keywords: Non-parametric item response theory; General Health Questionnaire; Psychological Distress

Corresponding Author: Dr Frank Doyle, PhD

Corresponding Author's Institution: Royal College of Surgeons in Ireland

First Author: Frank Doyle, PhD

Order of Authors: Frank Doyle, PhD; Roger Watson; Karen Morgan; Orla McBride

#### Abstract: Background

Invariant item ordering (IIO) is defined as the extent to which items have the same ordering (in terms of item difficulty/severity - i.e. demonstrating whether items are difficult [rare] or less difficult [common]) for each respondent who completes a scale. IIO is therefore crucial for establishing a scale hierarchy that is replicable across samples, but no research has demonstrated IIO in scales of psychological distress. We aimed to determine if a hierarchy of distress with IIO exists in a large general population sample who completed a scale measuring distress.

#### Methods

Data from 4107 participants who completed the 12-item General Health Questionnaire (GHQ-12) from the Northern Ireland Health and Social Wellbeing Survey 2005-6 were analysed. Mokken scaling was used to determine the dimensionality and hierarchy of the GHQ-12, and items were investigated for IIO.

#### Results

All items of the GHQ-12 formed a single, strong unidimensional scale ( $H=0.58$ ). IIO was found for six of the 12 items ( $H\text{-trans}=0.55$ ), and these symptoms reflected the following hierarchy: anhedonia, concentration, participation, coping, decision-making and worthlessness.

#### Limitations

The cross-sectional analysis needs replication.

#### Conclusions

The GHQ-12 showed a hierarchy of distress, but IIO is only demonstrated for six of the items, and the scale could therefore be shortened. Adopting brief, hierarchical scales with IIO may be beneficial in both a clinical and research context.

# A hierarchy of distress and invariant item ordering in the General Health Questionnaire-12

## Authors

<sup>1</sup>Doyle, F., <sup>2</sup>Watson, R., <sup>3</sup>Morgan, K., <sup>1</sup>McBride, O.

<sup>1</sup>Division of Population Health Sciences (Psychology)

Royal College of Surgeons in Ireland

123 St Stephen's Green

Dublin 2

Ireland

<sup>2</sup>Professor of Nursing,

School of Nursing & Midwifery,

The University of Sheffield,

Sheffield,

UK

<sup>3</sup>PU-RCSI School of Medicine

Block B & D Aras 1,

MAEPS Building, MARDI Complex,

Jalan MAEPS Perdana,

43400 Serdang Darul Ehsan

Malaysia

**Corresponding author**

Frank Doyle, PhD

Division of Population Health Sciences (Psychology)

Royal College of Surgeons in Ireland

123 St Stephen's Green

Dublin 2

Ireland

Email: [fdoyle4@csi.ie](mailto:fdoyle4@csi.ie)

Tel: + 353 1 4022718

Tel: + 353 1 4022764

# **A hierarchy of distress and invariant item ordering in the General Health Questionnaire-12**

## **Abstract**

### Background

Invariant item ordering (IIO) is defined as the extent to which items have the same ordering (in terms of item difficulty/severity – i.e. demonstrating whether items are difficult [rare] or less difficult [common]) for each respondent who completes a scale. IIO is therefore crucial for establishing a scale hierarchy that is replicable across samples, but no research has demonstrated IIO in scales of psychological distress. We aimed to determine if a hierarchy of distress with IIO exists in a large general population sample who completed a scale measuring distress.

### Methods

Data from 4107 participants who completed the 12-item General Health Questionnaire (GHQ-12) from the Northern Ireland Health and Social Wellbeing Survey 2005-6 were analysed. Mokken scaling was used to determine the dimensionality and hierarchy of the GHQ-12, and items were investigated for IIO.

### Results

All items of the GHQ-12 formed a single, strong unidimensional scale ( $H=0.58$ ). IIO was found for six of the 12 items ( $H\text{-trans}=0.55$ ), and these

symptoms reflected the following hierarchy: anhedonia, concentration, participation, coping, decision-making and worthlessness.

### Limitations

The cross-sectional analysis needs replication.

### Conclusions

The GHQ-12 showed a hierarchy of distress, but IIO is only demonstrated for six of the items, and the scale could therefore be shortened. Adopting brief, hierarchical scales with IIO may be beneficial in both a clinical and research context.

### Keywords

Non-parametric item response theory; General Health Questionnaire;  
Psychological Distress

Word count (excluding Table and References: **1742**)

## **Introduction**

Item response theory states that scale items can be ordered along levels of a latent trait – with item ‘difficulty’ demonstrating whether items are difficult/severe (rare) or less difficult (common) (Embretson and Reise, 2000). Furthermore, if one has endorsed an item that is difficult, then it is probable that that person has also endorsed other less difficult items. A hierarchy of symptoms can therefore be established. As an example, reported symptoms of fatigue and suicidality are both considered as part of the listed symptoms of major depressive disorder (American Psychiatric Association, 2000).

However, while symptoms of fatigue are common, and less serious, suicidality is rarer but obviously far more serious (Bertolote et al., 2005; Botega et al., 2005; De Leo et al., 2005; Kant et al., 2003; Lewis and Wessely, 1992). Such hierarchies may have intrinsic value for clinicians as they will more quickly suggest symptom severity, rather than relying on somewhat arbitrary scale threshold scores, for example, when screening for depression.

Indeed, recent research has indicated that psychological morbidity may have an inherent hierarchical form, indicating common but less severe distress to rarer but more intense levels. The General Health Questionnaire (GHQ) is a popular instrument for assessing psychological morbidity, and comes in several different versions (e.g. Goldberg, 1992; Goldberg and Hillier, 1979). Across all versions, higher total scores reflect higher levels of psychological morbidity, or distress. Watson et al. (2008) demonstrated, through the use of non-parametric item response theory (Mokken scaling), that a hierarchy of distress was evident in a community sample of over 6000 participants who

completed the 30-item version of the GHQ. Nine of the items formed a reliable and strong hierarchy. Furthermore, they found that the highest difficulty rating was for “Felt that life wasn’t worth living”, whereas the lowest difficulty was found for “Been (un)able to face up to your problems”, possibly indicating that the extreme feeling of hopelessness is commonly preceded by an inability to cope with problems. Similarly, in a sample of cardiac patients, Doyle et al. (2011) showed that depressive symptom items showed a hierarchical form reflecting prevalence: fatigue (41%–75%), depression (21%–38%) and hopelessness (10–11%). Finally, it is not just negative emotions that may be hierarchical, as recent findings suggest that happiness may also have such a structure (Stewart et al., 2010).

Although these studies suggest a hierarchy of emotions, it is possible that the ordering of items varies across samples, thus indicating that these scales are not true hierarchies, and rendering them less valuable (Ligtvoet et al., 2010). Invariant item ordering (IIO) is defined as the extent to which items have the same ordering (in terms of item difficulty) for each respondent who completes a scale (regardless of the individual’s total scale score) (Ligtvoet et al., 2011; Watson et al., in press). IIO is therefore crucial for establishing a scale hierarchy that is replicable across samples, and would provide powerful evidence for the utility and generalisation of such scales. Recently, a method for investigating IIO has been developed (Ligtvoet et al., 2010), and behavioural scales (e.g. physical function, feeding) have been found that demonstrate IIO (Watson et al., in press). However, it is unknown to what extent, if any, symptoms of distress hold the property of IIO. We therefore

analysed data from a representative **general** population survey to determine if a hierarchy of distress, with IIO, could be demonstrated.

## **Methods**

### NIHSWS

The Northern Ireland Health and Social Wellbeing Survey (NIHSWS; <http://www.csu.nisra.gov.uk/survey.asp5.htm>), 2005-2006, is a periodic survey of the health and wellbeing of the Northern Ireland population. It focuses on a range of different general health issues (including mental health) and health behaviours. We analysed data from 4107 of 4245 participants (**58.9% female, mean [SD] age 48.0 [18.1]**) who completed all items of the GHQ-12 (Goldberg, 1992). The data were obtained from the UK data archive (<http://www.data-archive.ac.uk/>). GHQ-12 items were scored in a Likert format, 1 “better than usual”, 2 “same as usual”, 3 “less than usual”, 4 “much less than usual”. Higher scores indicate higher levels of distress.

### Statistical analysis

#### Mokken scaling

Mokken scaling is an iterative scale-building technique (Meijer and Baneke, 2004). Mokken scaling is a probabilistic version of Guttman scaling – where a positive endorsement of one binary item of given difficulty indicates that remaining items of lesser difficulty have also been endorsed (**Guttman, 1950**) – but the Mokken procedure can also use polytomous items. Loevinger’s (**1948**) coefficient ( $H$ ) is used to interpret the results:  $H = [1 - (\text{observed Guttman errors}/\text{predicted Guttman errors})]$ . Expected Guttman errors are the



probability that the items are chosen by chance, while observed Guttman errors are the number of times items are endorsed as if not in an ordered sequence. Convention states that  $0.3 \geq H < 0.4$ ,  $0.4 \geq H < 0.5$  and  $H \geq 0.5$  indicate weak, moderate and strong scales respectively. Higher H values indicate higher item discrimination power, and thus more confidence in ordering of respondents (Meijer and Baneke, 2004). Following a recommended procedure, which involves increasing the co-efficient value until the most interpretable solution is found, items that demonstrate poor discriminability are excluded from the scale (Meijer and Baneke, 2004). Reported scales were found with a H-value set at 0.4, and these are ordered in terms of difficulty. Mokken scale analysis was conducted using a procedure written for Stata SE 9.2 (StataCorp, 2005), by Jean-Benoit Hardouin (Hardouin, 2004).

## IIO

Following Mokken scaling, two further steps were conducted to determine IIO: **the data were imported into the public domain statistical software R and using the Mokken Scaling Analysis programme the items forming the Mokken scale were analysed for IIO using the method *Manifest IIO* (van der Ark, 2007) which selects items with item response functions (IRFs) that do not overlap. Subsequently, these items were checked for accuracy in IIO using the statistic *H-trans*—analogous to Loevinger’s coefficient—which is a measure of the distance between the IRFs: the greater the distance the greater the accuracy of the IIO (Ligtvoet et al., 2010) with *H-trans* > 0.5 indicating a strong scale with respect to IIO.**

## Results

Mokken scaling found a single, strong unidimensional scale from the GHQ-12 items in the NIHSWS data at a threshold of  $H=.4$  (Table 1).

-----

Insert Table 1 about here

-----

No items were dropped from the analysis.

After seven steps of the *Manifest IIO* procedure, IIO was found for six of the 12 items, with a  $H\text{-trans}=0.55$ . These are indicated in Table 1 with asterisks.

## Discussion

We aimed to determine if a hierarchy of distress, with IIO, was evident in a general population sample who completed the GHQ-12. While a strong unidimensional scale was found, retaining all 12 items of the GHQ-12, only six of these items demonstrated IIO. **However, only one item was common with those found previously by Watson et al. (2008): “Been thinking of yourself as a worthless person”.** As women were under-represented in the Watson et al. sample, it is possible that the higher proportion of women seen in the present study may account for some of these differences.

The order of items is somewhat comparable to the findings of Watson et al. (2008), although perhaps does not have such a clear hierarchy of distress. For example, feelings of worthlessness were the most difficult item found here, whereas feelings of hopelessness were the most difficult in Watson et al. (2008). This finding is somewhat different to a recent study of the GHQ-12 in an Australian population survey which, using factor analysis, found a two-factor solution (O'Connor and Parslow, 2010). However, it should be noted that Mokken scaling is considered a superior test of dimensionality (Bech et al., 2011; Meijer and Baneke, 2004). Furthermore, increasing the H-value in our analyses omitted some items, but did not lead to another interpretable scale (data not shown). It may be that the GHQ performs differently in separate cultures.

Half of the items from the GHQ-12 showed IIO. IIO is not demonstrated when several items tap the same level of the latent trait (Ligtvoet et al., 2010; Watson et al., in press). This suggests that several of the items of the GHQ-12 assess the same level of the latent trait of distress, and this is supported by the fact that the mean score for five of these items was between 2.02–2.07. It is unsurprising that IIO was not found for these items, and such items may be redundant. This should be investigated in future work, and a reduction in the number of such items would strengthen the clinical and research applicability of the scale – **although it is possible that this may result in a scale that is more unreliable or may have an impact on construct validity. It should also be noted that the items which demonstrated IIO**

**are generally positively-worded, suggesting that the GHQ-12 may be better considered as a wellbeing scale, rather than a distress scale.**

Our results are limited by the fact that this was a cross-sectional analysis only.

Future research should try to replicate these findings to determine if similar items demonstrate IIO.

Table 1: Mokken scale of the GHQ-12 in the NIHSWS

<b>Item</b>	<b>Mean Score</b>	<b>H coefficient</b>
*11) have you recently been thinking of yourself as a worthless person?	1.42	0.62
10) have you recently been losing confidence in yourself?	1.67	0.65
6) Have you recently felt you couldn't overcome your difficulties?	1.76	0.62
9) Have you recently been feeling unhappy and depressed?	1.82	0.66
2) have you recently lost much sleep over worry?	1.91	0.56
*4) have you recently felt capable of making decisions about things?	2.02	0.48
5) have you recently felt under constant strain?	2.04	0.59
12) have you recently been feeling reasonably happy, all things considered?	2.06	0.57
*8) have you recently been able to face up to your problems?	2.07	0.58
*3) have you recently felt that you are playing a useful part in things?	2.07	0.41
*1) have you recently been able to concentrate on whatever you are doing?	2.16	0.50
*7) have you recently been able to enjoy your normal day-to-day activities?	2.18	0.54
Overall Scale		0.58

\*Items demonstrating IIO.

## References

- American Psychiatric Association, 2000. Diagnostic and statistical manual of mental disorders (text-revised), DSM-IV-TR. American Psychiatric Association, Washington, DC.
- Bech, P., Fava, M., Trivedi, M.H., Wisniewski, S.R., Rush, A.J., 2011. Factor structure and dimensionality of the two depression scales in STAR\*D using level 1 datasets. *J. Affect. Disord.* 132, 396-400.
- Bertolote, J.M., Fleischmann, A., De Leo, D., Bolhari, J., Botega, N., De Silva, D., Tran Thi Thanh, H., Phillips, M., Schlebusch, L., Varnik, A., Vijayakumar, L., Wasserman, D., 2005. Suicide attempts, plans, and ideation in culturally diverse sites: the WHO SUPRE-MISS community survey. *Psychol. Med.* 35, 1457-1465.
- Botega, N.J., Barros, M.B., Oliveira, H.B., Dalgalarrrondo, P., Marin-Leon, L., 2005. Suicidal behavior in the community: prevalence and factors associated with suicidal ideation. *Rev. Bras. Psiquiatr.* 27, 45-53.
- De Leo, D., Cerin, E., Spathonis, K., Burgis, S., 2005. Lifetime risk of suicide ideation and attempts in an Australian community: prevalence, suicidal process, and help-seeking behaviour. *J. Affect. Disord.* 86, 215-224.
- Doyle, F., Kowalczyk, A., McGee, H.M., Conroy, R.M., 2011. Exhaustion, depression and hopelessness in cardiac patients: A unidimensional hierarchy of symptoms revealed by Mokken scaling. *Ir. J Psychol. Med.* 28, 29-31.
- Embretson, S.E., Reise, S.P., 2000. Item response theory for psychologists. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Goldberg, D., 1992. General Health Questionnaire (GHQ-12). NFER-Nelson, Windsor, UK.

Goldberg, D.P., Hillier, V.F., 1979. A scaled version of the General Health Questionnaire. *Psychol. Med.* 9, 139-145.

**Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stoufer et al. (Eds.), *Measurement and prediction* (pp. 60-90). Princeton: Princeton University Press.**

Hardouin, J.-B., 2004. MSP: Stata module to perform the Mokken Scale Procedure. Boston College Department of Economics.

Kant, I.J., Bultmann, U., Schroer, K.A., Beurskens, A.J., Van Amelsvoort, L.G., Swaen, G.M., 2003. An epidemiological approach to study fatigue in the working population: the Maastricht Cohort Study. *Occup. Environ. Med.* 60 (Suppl 1), i32-39.

Lewis, G., Wessely, S., 1992. The epidemiology of fatigue: more questions than answers. *J. Epidemiol. Community. Health.* 46, 92-97.

Ligtvoet, R., van der Ark, L.A., Bergsma, W.P., Sijtsma, K., 2011. Polytomous Latent Scales for the Investigation of the Ordering of Items. *Psychometrika.* 76, 200-216.

Ligtvoet, R., van der Ark, L.A., te Marvelde, J.M., Sijtsma, K., 2010. Investigating an Invariant Item Ordering for Polytomously Scored Items. *Educ. Psychol. Measurement.* 70, 578-595.

**Loevinger, J., 1948. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychol. Bull.* 45,507-29.**

- Meijer, R.R., Baneke, J.J., 2004. Analyzing psychopathology items: a case for nonparametric item response theory modeling. *Psychol. Method.* 9, 354-368.
- O'Connor, D.W., Parslow, R.A., 2010. Mental health scales and psychiatric diagnoses: responses to GHQ-12, K-10 and CIDI across the lifespan. *J. Affect. Disord.* 121, 263-267.
- StataCorp, 2005. *Stata Statistical Software: Release 9.* StataCorp LP, College Station, TX.
- Stewart, M.E., Watson, R., Clark, A., Ebmeier, K.P., Deary, I.J., 2010. A hierarchy of happiness? Mokken scaling analysis of the Oxford Happiness Inventory. *Pers. Individ. Diff.* 48, 845-848.
- van der Ark, L.A., 2007. Mokken Scale Analysis in R. *J. Stat. Soft.* 20, 1-19.
- Watson, R., Deary, I.J., Shipley, B., 2008. A hierarchy of distress: Mokken scaling of the GHQ-30. *Psychol. Med.* 38, 575-579.
- Watson, R., van der Ark, L.A., Lin, L.C., Fieo, R., Deary, I.J., Meijer, R.R., in press. Item response theory: How Mokken scaling can be used in clinical practice. *J. Clin. Nurs.* doi: **10.1111/j.1365-2702.2011.03892.x**



## **Contributors**

FD wrote the first draft of the manuscript and analysed the data. RW analysed the data, and provided critical input. KM, OMcB provided critical input. All authors contributed to and have approved the final manuscript.

**Conflict of interest**

The authors have no conflicts of interest.

## Acknowledgements

For the use of the *Northern Ireland Health and Social Wellbeing Survey* data, the authors acknowledge the Northern Ireland Statistics and Research Agency, Central Survey Unit (sponsored by the Department of Health, Social Services and Public Safety (Northern Ireland)), and the UK Data Archive, University of Essex, Colchester. These bodies bear no responsibility for the further analysis or interpretation of this data.

**Role of Funding Source**

No funding was received for this study.