Personality and Individual Differences 94 (2016) 272–276

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Mokken scale analysis and confirmatory factor analysis of the Health of the Nation Outcome Scales

Steven J. Muncer [a,*], Barry Speak [b]

[a] Clinical Psychology, University of Teesside, Middlesbrough, UK
[b] Dept of Psychology Management, Tees, Esk and Wear Valley Trust, Middlesbrough, UK

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form 14 January 2016
Accepted 18 January 2016
Available online xxxx

Keywords:
HoNOS
Psychometrics
Outcome assessment
Mental health services

ABSTRACT

The Health of the Nation Outcome Scales is/are widely used in many countries and have been suggested for use within the Payment By Results framework in England. There are, however, questions over the best way to score it. Originally it was promoted as twelve independent scales, although more recently several alternative models of its factor structure have been proposed. Research so far has concentrated on the use of parametric methods of analysis that may be inappropriate. In this study, we examined the structure of 80,161 completed HoNOS scores using Mokken scale analysis, which is a nonparametric form of item response theory. Confirmatory factor analysis was also conducted on the proposed scales. Two possible subscales were found, which correspond to a Depression subscale and a Social and Cognitive Problems subscale. Neither scale had strong Mokken scale properties, particularly when compared with other scales. Confirmatory factor analysis suggested that the two subscale model had acceptable fit statistics. It is clear that the items cannot be considered as twelve independent items or indeed as a unidimensional scale. Given the relative psychometric weakness of HoNOS, it may be advisable to develop a new measure or at least to consider alternative measures of outcome.

1. Introduction

The Health of the Nation Outcome Scales is/are argued to consist of 12 clinician rated scales that measure problem severity for health and social functioning variables using a 0 to 4 point scale (Wing et al., 1998). They are widely used in Britain and also in Australia, New Zealand and Canada (Kisely, Campbell, Cartwright, Cox, & Campbell, 2010; Kisely, Xiao, Crowe, Paydar, & Jian, 2014; Page, Hooke, & Rutherford, 2001) and also have been used in Europe (Lovaglio & Monzani, 2012). There is, however, some ambiguity as to whether they are a set of twelve independent scales as was stated as the original intention, or a scale with twelve related items so that they can be added to produce a meaningful total score (Hampson, Killaspy, Mynors-Wallis, & Meier, 2011; Speak, Hay, & Muncer, 2015). One of the reasons for this argument has been the difficulty of getting an agreed factor structure for HoNOS and, therefore, a template for aggregating item scores. There have been at least six factor structures proposed for HoNOS, which are shown in Table 1. Studies of the structure of HoNOS have tended to start by showing that Wing et al's (1998) model is not a good fit, and then proposing an alternative. Some of these (McClelland, Trimble, Fox, Stevenson, & Bell, 2000; Newnham, Harwood, & Page, 2009; Preston, 2000; Speak et al., 2015; Trauer, 1999) have tried to fit all 12 items into a factor structure, whereas Lovaglio and Monzani (2012) proposed a one factor structure with only 6 items. All of the proposed structures also include correlations between item residuals or other complications to the model which are not indicated in the table. It would be fair to say, however, that none of the models has consistently good fit statistics when subjected to confirmatory factor analysis, although the Speak et al. (2015) model has been shown to be better than the other twelve item models.

It should be noted that there is no evidence that the 12 HoNOS items can be considered as independent scales, as in every study of a proposed model the independence model is significantly worse than the model that was proposed (Williams, Speak, Hay, & Muncer, 2014). The failure to find a satisfactory consistent structure cannot be taken in itself as evidence for independence of the twelve items (Andreas et al., 2010). It probably means that some of the items do not fit any structure hence Lovaglio and Monzani's (2012) attempt to look at a selection of items. Alternatively it could mean that no satisfactory structure of the twelve items has been proposed so far.

Classical test theory is predicated on item correlations that test the extent to which people respond similarly to items intended to measure the same trait. Its ability to answer questions regarding the dimensionality of some scales has been argued to be poor (van der Linden & Hambleton, 1997; van Schuur, 2003). Furthermore factor analytic approaches rely on strong assumptions about the level of the data and its distribution. Here is one of the problems with using statistical methods based on classical test theory for any analysis of HoNOS.

* Corresponding author at: Clinical Psychology, Teesside University, Middlesbrough TS1 3BA, UK.
E-mail address: s.muncer@tees.ac.uk (S.J. Muncer).

http://dx.doi.org/10.1016/j.paid.2016.01.051
0191-8869/

Table 1
Previously proposed factor structures of HoNOS.

Item	Wing/Preston	Newnham	McLelland	Trauer	Speak	Lovaglio
1 Behaviour disturbance	1	1	1	1	1	
2 Self-injury	1	1	2	4	2	
3 Drinking/drug use	1	1	1	1	3	
4 Cognitive problems	2	2	1	2	4	1
5 Physical illness	2	2	3	2	4	
6 Hallucinations/beliefs	3	2	1	3	1	1
7 Depressive symptoms	3	3	2	4	2	
8 Other mental health problems	3	3	4	4	2	
9 Social relations	4	1	1	4/5	3	1
10 Activities of daily living	4	2	1	5	4	1
11 Living conditions	4	4	1	5	3	1
12 Daytime activities	4	4	1	5	3/4	1

Responses to the items are not normally distributed and neither are they multivariate normal, which are requirements for confirmatory factor analysis. It is possible that some of the confusion over the structure of HoNOS has been caused by mainly using forms of factor analysis, or at least not considering alternative methods of analysis.

Increasingly, researchers are moving beyond classical test theory in evaluating the properties of psychometric and clinical measures. Item response theory examines the structure of the items based on ordering them in terms of difficulty (Mokken, 1971). It is based on the premise that an individual who achieves a high overall score would be more likely to get a difficult question correct than someone who gets a lower overall score. This can also be applied to personality and clinical measures to reveal the hierarchical structure of the items.

Mokken scaling is based on item-response theory and is similar to Rasch scaling techniques but has the advantage of having fewer restrictions in its use (Mokken, 1971; Meijer, Sijtsma, & Smid, 1990; Stochl, Jones, & Croudace, 2012). Although based on Guttman scaling, Mokken does not assume error-free data. Nor does it include assumptions about the sigmoid shape of item characteristic curves that can result in the rejection of many items and so decrease the reliability of the resultant measure. In the present study we first used exploratory Mokken scale analysis to examine the scale structure of HoNOS. The scales that were developed were then tested on a new large sample to confirm their existence, again using Mokken scale analysis. Lastly, CFA was conducted on the model that was developed through these Mokken scale analyses.

2. Method

2.1. Participants

The national sample consisted of 80,161 HoNOS ratings taken at the point of referral. These were from patients who were referred to working age adult and older persons' mental health services from across the whole of England. The records were extracted from the MHMDS by the Health and Social Care Information Centre (HSCIC), anonymised and collated to form a national database (Copyright © 2013, Reused with permission of the Health & Social Care Information Centre. All rights reserved). Data quality checks were performed prior to analysis and only records with complete HoNOS scores were used. This made any analysis of participants with missing responses impossible. It should be noted, however, that the sample is well over the minimum sample size requirements for Mokken scale analysis (Straat, van der Ark, & Sijtsma, 2014). It would have been useful, however, to examine missing data to look for patterns of missing items and also possible causes of noncompletion of items. Demographic information was recorded for 95.8% of the sample. 50.4% were female and 45.4% were male. The average age was 49.9 years (SD = 20.4 years). 32.2% of the sample were patients with non-psychotic disorders; 32.6% to patients with psychotic disorders; and 17.6% to patients with organic disorders. Diagnosis was not recorded for 17.6% of the sample. The sample was randomly divided into two groups so that any useful models could be retested on a new sample.

2.2. Statistical analysis

The most important calculation in Mokken scale analysis is the calculation of Loevinger's coefficient H, which is based on the extent to which item pairs conform to Guttman scaling. For a Guttman scale pairs of items should have a consistent relative score to each other so if one item is considered as more difficult or less likely to be endorsed than another, this pattern should be consistent across participants. If a participant endorses the items in the wrong direction, so that the easier to be endorsed item is endorsed less than the hard to endorse item, then this is an error. Loevinger's H calculates the size of this error for each item, each pair of items and also the overall scale.

A Mokken scale for polytomous items must first meet the criteria of the monotone homogeneity model, which is defined by three assumptions. Firstly, the data should be unidimensional (items assess the same latent trait). Second, it should show monotonicity (the probability of any given response is a non-decreasing function of that trait).This means that higher latent trait values increase the probability of correct responses to items measuring that latent trait. Third, items should show local independence (participants' response to any given item is not influenced by their response to other items). It is also desirable that the items show invariant item ordering (IIO) (Ligtvoet, van der Ark, Te Marvelde, & Sijtsma, 2010), which means that items can be ordered according to their difficulty (or frequency of endorsement), allowing for hierarchical ordering of scale items. This requires the calculation of three coefficients. Coefficient H*i* for each item provides a measure of the item's scalability (and unidimensionality). From these values, an H coefficient for the full scale can be calculated which indexes the extent to which scale items accurately order respondents. Htrans (H^T) reverses the role of persons and items, and thus indexes the extent to which a sample of individuals agree on the ordering of the items (Sijtsma, Meijer, & van der Ark, 2011). Taken together, H and H^T are indicative of the strength and structure of a scale. Ultimately, if the DMM fits the data, and IIO can be established, it can be concluded that item ordering is robust across populations and population sub-groups (Sijtsma et al., 2011). Mokken scale analysis also produces a reliability of the scale statistic rho, which is similar to Cronbach's alpha, acceptable alpha is also taken as >.7 (DeJong & Molenaar, 1987).

First exploratory Mokken scale analysis (Mokken, 1971) was used to investigate the scalability and dimensionality of HoNOS. The 'Mokken' package in R was used for this purpose (van der Ark, 2007). The exploratory analysis followed Hemker's procedure (Hemker, Sijtsma, & Molenaar, 1995) in which the automated item selection procedure is used to select items to form scales. An iterative process is followed in which the lowerbound cut off (c) for acceptable item H values starts at 0 and is raised in 0.05 increments until 0.6. Scales developed through