A hierarchy of happiness? Mokken scaling analysis of the Oxford Happiness Inventory

Mary E. Stewart [a,*], Roger Watson [b], Andrea Clark [a], Klaus P. Ebmeier [c], Ian J. Deary [d]

[a] Applied Psychology, School of Life Sciences, Heriot-Watt University, Edinburgh, United Kingdom
[b] School of Nursing, Midwifery and Health, The University of Sheffield, United Kingdom
[c] Department of Psychiatry, University of Oxford, United Kingdom
[d] Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, United Kingdom

ARTICLE INFO

ABSTRACT

The items of the Oxford Happiness Inventory (OHI), a self-report assessment of happiness, are subjected to an analysis for hierarchy among its items. By using Mokken scaling analyses we can assess whether items can reliably be ordered between persons as severity indicators on a latent trait; in this case, a latent trait of Happiness. OHI item-level data from 1024 participants were entered into the Mokken Scaling Procedure (MSP) seeking reliable scales with $H > 0.30$. 12 OHI items formed a reliable and statistically significant hierarchy. However, the MSP values indicate a 'weak' scale. The 'most difficult' (happiest) item on the scale is 'feeling energetic' and the 'least difficult' (least happy) is 'I have fun'. Items in the scale are consistent with what is already known about both happiness and low mood. The reduction in the OHI's items from 29 to 12 in the Mokken scale may have utility making it more accessible to participants as well as identifying items with reliably different levels of 'difficulty'.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Research into happiness has increased considerably over the past forty years. A product of this interest in happiness, its causes, and the psychological processes that produce it, has been the development of several self-report scales which aim to measure this construct (Kashdan, 2004; Larsen, Diener, & Emmons, 1985). These subjective measures have allowed progress in happiness research by supplementing early social surveys based on objective variables (Swami, 2008) or 'social' indicators' (Argyle, 2001) of happiness, such as health, income, education, demographics and life events. They have helped to provide a fuller account of happiness (Diener, 1984).

Happiness is generally considered to comprise three main components: the frequency and degree of positive affect or joy; the absence of negative feelings, such as depression or anxiety; and the average level of satisfaction over a period (Andrews & Withey, 1976; Diener, 1984). The first two refer to the affective aspects of the construct, the latter to the cognitive aspects (Diener, Emmons, Larsen, & Griffin, 1985). Happiness is, therefore, a multidimensional construct. Some measures of happiness emphasise the affective component (e.g., the Affect Balance Scale, Bradburn, 1969; the

Positive and Negative Affect Scale, Watson, Clark, & Tellegen, 1988), some emphasise the cognitive component (e.g., the Life Satisfaction Scale, Diener et al., 1985; the Delighted-Terrible Scale, Andrews & Withey, 1976), and others offer a combination as general measures of happiness (e.g., the Oxford Happiness Inventory Argyle, Martin, & Crossland, 1989).

Scales measuring happiness range from single questions to multi-item scales. In early research on happiness, researchers tended to rely on a single self-report item to measure each component of happiness (Diener, 2000). Single-item scales, for example, "taking all things together, how would you say things are these days?" (Gurin, Veroff, & Feld, 1960), have been rated using Likert scales, verbal rating scales (e.g., 'very happy', 'pretty happy', 'not too happy'), or by visual analogue scales. Cantril's (1985) 'ladder scale' and Andrews and Withey's (1976) Delighted-Terrible Scale use seven schematic faces whose expressions range along the continuum from very negative to very positive. Although single-item measures can be quite successful for some research investigations (Argyle, 2001), such measures have been criticised in that they cannot be tested for internal consistency reliability (Argyle, 2001; Diener, 1984; Stones et al., 1996).

Several multi-item measures exist and those that have been widely used include Bradburn's (1969) Affect Balance Scale, the Depression-Happiness Scale (McGreal & Joseph, 1993), Satisfaction With Life Scale (Diener et al., 1985), the Memorial University of

Newfoundland Scale of Happiness (Kozma & Stones, 1980; Kozma & Stones, 1983), the Short Happiness and Affect Research Protocol (Stones et al., 1996), the Subjective Happiness Scale (Lyubomirsky & Lepper, 1999), and the Oxford Happiness Inventory (OHI: Argyle et al., 1989).

The present study will assess the hierarchical properties of the items in the OHI, using Mokken scaling. As far as we are aware the OHI has not been subjected to any analysis searching for a hierarchy among its items. A scale with hierarchical properties has items which are consistently ordered relative to one another, giving an indication of the relative position of each item on the latent trait assessed by the scale; the extent to which each item indicates a 'mild' or 'severe' level of the latent trait. To explain this psychometric concept, consider the following analogy. If weight lifters can successfully lift 130 kg, they would not be asked to lift 120 or 110 kg because these weights are easier. Therefore, if, on a scale, a person endorses an item indicating a given level of happiness, they are likely also to have endorsed all items indicating lower levels of the same latent trait. However, this cannot be taken for granted, and the hierarchy of items can be tested empirically.

Hierarchical scales have proven useful in measuring a range of constructs: for instance, Neuroticism (Watson, Deary, & Austin, 2007), psychological distress (Watson, Deary, & Shipley 2008) and feeding behaviour in dementia (Watson, 1996). Such scales enable the presence of a latent trait to be determined, not by finding items' loadings as is done in factor analysis, but by discovering whether and how many specific items form an ordered array along it. The General Health Questionnaire (GHQ) has several formats with 60, 30, 28 or 12 items. The GHQ is often used to measure psychological distress and levels of mental ill-health. It was originally developed to assess the extent of psychiatric illness in general practice (Goldberg, 1978). Watson et al. (2008) identified nine items that form a useful and reliable scale by applying Mokken scaling to the GHQ-30. A shorter scale when assessing mental health is of particular clinical utility as it is often difficult for participants to complete long questionnaires.

The Oxford Happiness Inventory (OHI; Argyle et al., 1989) was designed to provide a general measure of personal happiness and followed the design of the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), with several of the BDI items reversed and, in addition, supplementary items measuring further aspects of subjective well-being (Argyle, 2001). The 29-item OHI follows the same four-choice format as the BDI, with incremental steps defined as follows: unhappy or mildly depressed (e.g., 'I do not feel happy'); a low level of happiness (e.g., 'I feel fairly happy'); a high level of happiness (e.g., 'I am very happy'); and manic (e.g., 'I am incredibly happy') (Francis, 1999). The aggregate score forms the measure of overall happiness, with higher scores indicating greater happiness. Cronbach alpha coefficients for the OHI are high (0.84–0.90). However, despite these high alphas for the total scale, factor analysis identifies, in addition to a single factor of happiness, several more specific dimensions: satisfaction with life; personal efficacy; sociability/empathy; a positive outlook; physical well-being; cheerfulness; and self-esteem (Argyle, Martin, & Lu, 1995; Hills & Argyle, 1998).

The OHI has been widely used in the UK (e.g., Furnham & Brewin, 1990; Joseph & Lewis, 1998) and USA (e.g., Valiant, 1993), and has been used cross-culturally to compare students in Australia, Canada, the UK and USA (Francis, Brown, Lester, & Philipchalk, 1998). There are also several translations of the OHI (see Lewis, Francis, & Ziebertz, 2002). The scale shows high internal reliability (e.g., Argyle & Lu, 1990; Hills & Argyle, 2001), with high reliability over time (Argyle et al., 1989; Francis, Ziebertz, & Lewis, 2003), and good consensual validity when self-reports are compared with a friend's ratings (Argyle, 2001). It also correlates with measures of all three components of happiness, with large correlations between

the OHI and life satisfaction, and with depression: Positive affect (Bradburn positive affect; .32); life satisfaction (Life Satisfaction Index; .57); negative affect (Bradburn negative affect; −.32: BDI; −.52) (Argyle et al., 1989).

In the present study, we inquire whether items from the OHI form a hierarchy by using Mokken scaling. We test whether the items can reliably be ordered as severity indicators on a latent trait. If OHI items form a Mokken scale it will help to understand the latent trait of Happiness.

2. Methods

Mokken scaling, derived from one of the two branches of item response theory (Hulin, Drasgow, & Parsons, 1983, p. 14), is a stochastic development of Guttman scaling (van Schur, 2003). Both methods seek to establish whether hierarchical scales exist in multivariate datasets. However, whereas Guttman scaling can only process dichotomous responses to items in a questionnaire, Mokken scaling can process multiple response categories. This is enabled through a computer programme: the Mokken Scaling Procedure (MSP; Sijtsma, Debets, & Molenaar, 1990). Furthermore, whereas Guttman scaling is deterministic in nature and very prone to violations by items of the Guttman hierarchy, Mokken scaling, due to its stochastic nature, is more flexible in this respect (Mokken & Lewis, 1982).

Mokken scaling generates the following informative parameters: *scalability* (Loevinger's coefficient or H), *reliability* (Rho; a test–retest statistic analogous to Cronbach's alpha), and *probability* (p, Bonferroni corrected for multiple comparisons) of obtaining a hierarchical scale (Molenaar & Sijtsma, 2000). *Scalability*, in Mokken scaling terms, refers to the extent to which items in a questionnaire will be ordered hierarchically relative to one another. The Mokken scaling procedure uses the mean scores for endorsement of items to order the items. In item response theory, items in a hierarchy are described as being ordered from 'easy' to 'difficult' which refers to the extent to which they are endorsed by respondents. Where a high score on an item indicates endorsement, 'easy' items are readily endorsed and, therefore, have high mean scores and 'difficult' items are less readily endorsed, thereby having lower mean scores. The Loevinger's coefficient is calculated from a set of items ordered, on the one hand, according to their mean values and, on the other hand, taking into account the number of Guttman violations in the relative ordering of items; the fewer Guttman violations, the greater the scalability. Acceptable values for these criteria have been established as $H > 0.3$; $Rho > 0.7$, and $p < 0.05$.

3. Procedure

3.1. Recruitment of participants

Participants were recruited from Edinburgh's Universities. They were included if English was their first language and if they were over 17 years of age. Participants were given a stamped, addressed envelope to return a questionnaire, and were asked to take a questionnaire only if they were willing to be considered for a later, experimental (acute tryptophan depletion) stage of the study (see Stewart, Deary, & Ebmeier, 2002). Approximately 2000 questionnaires assessing mood and personality were distributed; 1041 questionnaire packs were completed and returned, among which 1036 completed the OHI. The questionnaire pack took approximately 1 h to complete. Data concerning the following were described previously: the structure of the TPQ (Stewart, Ebmeier, & Deary, 2004); the relationships between personality and mood (Stewart, Ebmeier, & Deary, 2005), and whether personality is a