Does cognitive ability influence responses to the Warwick-Edinburgh Mental

Well-being Scale?

Abstract

It has been suggested that how individuals respond to self-report items relies on cognitive processing. We hypothesised that an individual's level of cognitive ability may influence these processes such that, if there is a hierarchy of items within a particular questionnaire, as demonstrated by Mokken scaling, the strength of that hierarchy will vary according to cognitive ability. Using data on 8643 men and women from the National Child Development Survey (1958 Birth Cohort), we investigated, using Mokken scaling, whether the 14 items that make up the Warwick-Edinburgh Mental Well-being Scale—completed when the participants were aged 50 years—form a hierarchy, and whether that hierarchy varied according to cognitive ability at age 11 years. Among the sample as a whole, we found a moderately strong unidimensional hierarchy of items (Loevinger's coefficient (H)=0.48). We split participants into three groups according to cognitive ability and analysed the Mokken scaling properties of each group. Only the medium and high cognitive ability groups had acceptable ($\geq$0.3) invariant item ordering (IIO, assessed using the H^T statistic). This pattern was also found when the three cognitive ability groups were assessed within men and women separately. Greater attention should be paid to the content validity of questionnaires to ensure they are applicable across the spectrum of mental ability.

Key words: Mokken scaling, hierarchical scales, item response theory, cognitive ability, mental well-being

Does cognitive ability influence responses to the Warwick-Edinburgh Mental Well-being Scale?

According to the World Health Organization, mental health is more than the absence of mental disorders, but is a "state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" (Herrman, Saxena & Moodie, 2005). To understand individual differences in well-being and their determinants, it is important to have reliable and valid measures. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS) was developed by an expert panel in response to increasing recognition that, if we are to have a full picture of the levels of mental health in a population and understand the factors that influence it, there is a need for measures of positive mental health to supplement the many instruments that assess symptoms of anxiety or depression, in other words the negative aspects of mental health (Huppert & Whittington 2003; Hu, Stewart-Brown, Twigg, & Weich, 2007).

The WEMWBS is potentially especially valuable because it is a measure of mental well-being that focuses entirely on positive aspects of mental health (Tennant et al., 2007). It has been used in national surveys of mental well-being in Scotland since 2006 (Corbett et al., 2010). The 14-item WEMWBS was designed to cover a broad concept of mental well-being, including affective or emotional aspects, cognitive or evaluative aspects, and psychological functioning. Individuals completing the scale are asked to tick the box that best describes their experience of each of the 14 statements over the past two weeks using a 5-point Likert scale. The total score indicates the level of mental well-being, with higher scores indicating

greater well-being. Confirmatory factor analysis suggests that the WEMWBS is measuring a single underlying concept (Tennant et al., 2007).

More recently Stewart-Brown et al. (2009) examined the internal construct validity of the scale according to the perspective of the Rasch Measurement Model. They found that some of the 14 items showed bias for gender (for example, at any level of well-being, men were more likely than women to report a higher score for the item 'I've been feeling confident'), and one item showed bias for age. In view of this, they suggest a 7-item version of the scale would have more robust measurement qualities, and this short version is now available. However, these authors also suggest that there are arguments for continuing to collect data on all 14 items so that item bias can be explored in different samples.

To our knowledge, there has been no investigation as yet into the WEMWBS using Mokken scaling. Mokken scaling is a method of analysing items within questionnaires or other instruments for the existence of cumulative hierarchical scales. In a Mokken scale the ordering of items relates the items specifically to levels of the latent trait while excluding items which do not meet the criteria of Mokken scaling. In this way, a shorter—but robust— scale could be produced which could, for example, be useful for screening purposes. Mokken scaling is based on item response theory; unlike Rasch analysis, it is non-parametric and, therefore, less restrictive (Gillespie, Tenvergert, & Kingma, 1988). Mokken scaling has proved useful in the analysis of a wide range of constructs, for example feeding behaviour in people with dementia and quality of palliative care (Ringdall, Jordhoy & Kaasa, 2003; Watson, 1996). It has also been used with psychological constructs, including neuroticism, happiness, and psychological distress (Stewart, Watson, Clark, Ebmeier & Deary, 2010; Watson, Deary & Austin, 2007; Watson, Deary & Shipley, 2008). Mokken

scaling analysis provides quantitative parameters to indicate whether items form a hierarchy: that is, whether the items in a scale are answered such that some items strongly tend to be endorsed before others by all respondents. This gives the notion of item difficulty and Mokken scaling can find out whether, for all individuals, the items have the same order of difficulty.

In a good Mokken scale the presence of the latent trait can be represented by the score on a single item—the highest one endorsed by respondents. Therefore, the first aim of the present study was to investigate the WEMWBS to discover whether the items had a hierarchy of endorsement in the subjects studied.

Most of the attention in Mokken scaling has been on the items, and asking whether or not they form a hierarchy because of how they are worded. Here, we shall raise an additional important issue and ask: might the Mokken hierarchy depend also on individual differences in people's ability to interpret the items?

The interpretation and response to any given item may be multifaceted. Karabenick et al. (2007), building on prior work by Hastie (1987) and Sudman, Bradburn & Schwart (1996), proposed a cognitive processing model of self-report items. Here, individual responses rely on an individual's ability to: a) read and interpret the meaning of words in an item; b) interpret the meaning of the item and store this in working memory; c) search memory for personal information relevant to the meaning; d) read and interpret the response format of the item; e) simultaneously evaluate the item word meanings, memory, and item response scale; and f) select the most congruent response option (Karabenick et al., 2007, p.141). A process such as that proposed by Karabenick et al. (2007), or any such analogous cognitive model, is clearly cognitively complex and demanding and, as such, individuals of greater cognitive ability may be more adept. There is some evidence to support this. For

example, reading comprehension, crucial for steps (a), (d) and (e) above, is positively associated with general cognitive ability (Johnson, Bouchard, Segal, & Samuels, 2005).

We hypothesise that an individual's level of cognitive ability may influence these cognitive processes such that the Mokken scale properties of groups at varying levels of cognitive ability may differ. If we consider the possibility that one group of people could understand the difference between two items' wordings and another group could not, then only the former group would afford the possibility of there being a consistent hierarchical ordering of those items. People of higher cognitive ability may not only have a better understanding of the meaning of words and phrases used in the items of a scale, but they may also be more accurate at judging how items may differ in terms of how objectively 'mild' or 'severe' they are on the underlying trait that they represent.

Few studies have looked specifically at whether respondent's levels of cognitive ability influence the scaling properties of individual constructs. At a structural level, researchers have considered the personality differentiation hypothesis (Brand, 1994), the concept that the structure of personality as a whole may differ across levels of cognitive ability (e.g. De Fruyt, Aluja, Garcia, Rolland, & Jung, 2006; Mottus, Allik, & Pullman, 2007; Rammstedt, Goldberg, & Borg, 2010). However, in general, differentiation studies say nothing of the individual scale properties, though some indirect support may be gleaned from aspects of differentiation studies. For example, lower internal consistencies have been noted for personality scales in lower IQ groups (Austin, Deary, & Gibson, 1997; Allik & McCrae, 2004). Further, Austin et al. (2002) suggested that high correlations observed between Psychoticism and

Neuroticism scores in a low IQ group, may in part be due to lower IQ respondents failing to differentiate items from different scales.

Waiyavutti, Johnson and Deary (2011) conducted a more comprehensive study, testing differential item functioning across cognitive ability groups. The authors conducted IRT and invariance analysis on the items of the NEO-FFI in a sample of 640 older adults (n=320 lower IQ; n=320 higher IQ). They found no statistically significant evidence for differential item function across levels of cognitive ability. However, the authors note a number of trends in responses to individual items, such as the endorsement of extreme ends of scales and acquiescence, particularly in the lower IQ group. In the case of the NEO-FFI, the extremes of responding resulted in a need to collapse Likert categories. Therefore, although no statistical differences in item functioning were found, there was moderate evidence of varied item performance across IQ levels, suggesting that further research may be justified.

The aims of the present study are to investigate, using Mokken scaling: whether the 14 items that make up the WEMWBS form a hierarchy; and whether the strength of that hierarchy varies in strength according to people's cognitive ability.

Methods

Participants

The National Child Development Study (1958 cohort) was originally based on 18558 births in Great Britain in one week in 1958 (Power & Elliott, 2006). The cohort has subsequently been followed-up at regular intervals. In total, 9790 study members took part in the 2008-2009 follow-up survey when they were aged 50 years, and during this survey 8643 (70%) completed the WEMWBS. Ethical approval for this

study was obtained from the South East Multicentre Research Ethics Committee. Of these, 7510 had taken a test of cognitive ability at age 11 years (Figure 1).

Cognitive ability

Cognitive ability was assessed at school when the children were aged 11 years using a general cognitive ability test, devised by the National Foundation for Educational Research in England and Wales (Douglas, 1964). The test consisted of 40 verbal and 40 non-verbal items and was administered by teachers. Total scores from this test correlate strongly with scores on a test of verbal ability used to select 11-year-old children for secondary school (r=0.93) suggesting a high degree of validity (Douglas, 1964).

Mokken scaling

The properties of a Mokken scale can be estimated using the model of monotone homogeneity (MMH) and invariant item ordering (IIO). MMH consists of three assumptions: unidimensionality, local stochastic independence, and monotonicity. Monotonicity means that item response functions are monotonously increasing. IIO consists of a single assumption: the non-intersection of item response functions. The assumption of IIO is, as described by Ligtvoet *et al*. (2010, p 593): "both omnipresent and implicit in the application of many tests, questionnaires, and inventories". It means that the ordering of items at the group level by mean scale scores also holds at the individual level. MMH is tested using Loevinger's coefficient (H) for individual items, pairs of items and the overall scale. H is a measure of the ratio of the observed to expected errors in the order or scalability of items which ranges from 0 (no scalability) to 1 (perfect scalability); $H > 0.3$ is the minimum value for an acceptable Mokken scale and items with $H < 0.3$ are removed from the analysis to produce an acceptable Mokken scale. IIO is tested using H^T (H-trans, analogous to

H, where H^T is a measure of the ratio of observed to expected violations of IIO) and scales with $H^T > 0.3$ are considered to show acceptable IIO. The probability of obtaining a Mokken scale can be estimated by a Bonferroni corrected method and also the reliability (Rho) of any scale by a method analogous to test-retest reliability.

Data were entered into an SPSS database and then converted to *.Rdata files and analysed using the Mokken scale analysis (MSA) facility in the R statistical package version 2.11.1 (van der Ark, 2007). SPSS data were also saved in tab delimited format with the spreadsheet option turned off and imported into the Mokken Scaling Procedure (MSP) for Windows (Molenaar & Sijtsma 2000). Using MSA in R, the data were analysed for IIO.

Initially, the complete dataset (n=8643) was analysed using the MSP to explore the possibility of a unidimensional hierarchy of items. IIO was not explored in the complete dataset due to limitations regarding sample size in the R programme. We then grouped participants into 3 groups according whether they had low (>1 SD < mean; n=857), medium (mean ± 1 SD; n=4671) or high (>1 SD > mean; n=1531) cognitive ability in childhood. After this, we divided the participants with medium cognitive ability into 2 groups according to whether they were male (n=2230) or female (n=2443) participants and then divided these into low, medium or high mental ability. The Mokken scaling properties of each group were analysed.

Results

Characteristics of the sample are shown in Table 1. An independent t-test showed that there was a significant difference in mental ability between males and females (mean difference 2.05; p<0.0001). The results of Mokken scale analyses are shown in Table 2. A moderately strong unidimensional hierarchy of items is shown under the model of MHH (H>0.40) except for the females of medium and high mental

ability for which a strong ($H>0.50$) hierarchy of items is shown. Acceptable IIO ($H^T \geq 0.30$) is shown for all except the low cognitive ability participants in the total sample and for both low cognitive ability males and low cognitive ability females for which it was considered too weak ($H^T<0.30$) for these to form a hierarchical scale. Generally, the hierarchy of items runs, in terms of 'difficulty'—indicated by the items' mean scores—from items related to clarity of thinking (I've been able to make up my own mind about things; I've been thinking clearly; I've been interested in new things) to stronger feelings of well-being (I've been feeling relaxed; I've been feeling optimistic about the future; I've energy to spare). Therefore, with regard to the first study aim, the WEMWBS does have a hierarchy of items.

In all of the scales, the ordering of items is broadly similar. One noticeable difference between the scales for males and females were items 4 (I've been feeling interested in other people) and 9 (I've been feeling close to other people) which were the third and fourth most endorsed item by female participants but which were fourteenth and twelfth, respectively, most endorsed by males participants.

In terms of IIO, items 4 (I've been feeling interested in other people), 8 (I've been feeling good about myself), and 14 (I've been feeling cheerful) only show IIO in one scale each, and this is not consistent across the sub-groups of the analysis. Item 10 (I've been feeling confident) does not show IIO in any sub-group.

Discussion

The study's first aim was to discover whether the WEMWBS showed a hierarchy of items. It does, whether this is for all subjects, for medium and high ability subjects, or for men and women. The second aim was to test the hypothesis that people with lower cognitive ability might have a less strong hierarchy of items, by our reasoning that completing the WEMWBS is in part a verbal cognitive task that

includes discriminating meaning differences between items and weighting them to some underlying construct for severity. This proved correct; the lower ability tertile—whether this was based on the whole sample, or within men or women—was the only group to have unacceptable IIO values..

Stewart-Brown et al. (2009) identified 7 items from the WEMWBS meeting the criteria for Rasch analysis that were strictly unidimensional. As they explained, few scales have been shown to meet the strict criteria of Rasch analysis; however, as can be deduced from the fact that the original WEMWBS contained 14 items, this is at the considerable expense of items in the scale. Higher scores on items retained in Rasch analysis are clearly related to higher levels of the latent trait, which is a very useful property to ascertain. The application of Mokken scaling has provided scales with greater numbers of items retained and, therefore, a more 'authentic' assessment of the latent trait as opposed to the very 'direct' assessment (Messick 1994) offered by the 7 items in the Rasch scale.

In addition, generally, to retaining more items, Mokken scaling has provided an ordering of items which relates items specifically to levels of the latent trait. Items not meeting the criteria of Mokken scaling were excluded. Therefore, Mokken scaling has demonstrated that most of the items in the WEMWBS are suitable for measuring the latent trait of mental well-being; however, caution must be exercised with people of low mental ability for whom IIO was considered too weak to indicate an hierarchical scale. On the other hand, for people with medium and high mental ability—and for men and women from the whole sample—items showed at least weak IIO. The implication is that people with medium to high mental ability can better interpret the meaning of the items in the WEMWBS and use the scoring system to indicate their mental well-being. On the other hand, presumably, the WEMWBS,

contains insufficient items properly to capture the mental well-being of people with low mental ability.

The strengths of this study lie in the availability of a large and generalisable sample that has enabled testing of responses to the WEMWBS according to different mental ability strata and gender. Adequacy of sample size in Mokken scaling is hard to estimate due to the ordinal nature of the scales, but it has been stated that Mokken scaling can be applied safely to samples of several hundred (Meijer & Baneke 2004). Therefore, for the present study, all the scales have been tested on adequate samples. This theory that mental ability may influence responses to psychological assessment instruments has therefore been tested and demonstrated for the WEMWBS. Clearly, wider testing across cultures and with other questionnaires needs to take place to investigate the transferability of our observations concerning cognitive ability and responses to self-report scales.

Our study has several implications. It has demonstrated the utility of Mokken scaling using the WEMWBS, showing that the resulting scales are more economical than the original scale and have the added value of relating items to levels of the latent trait. These Mokken scales, therefore, could be used for screening purposes to decide which individuals require fuller assessment. Further research into the relationship between mental ability and responses to questionnaires could be undertaken. Finally, in the design of future questionnaires for psychological assessment it should be taken into account that such questionnaires, generally designed by people with high mental ability, may be less suitable for use with people of lower mental ability. Thus greater attention should be paid to the content validity of questionnaires to ensure they have wider applicability across the spectrum of mental ability.

References

Allik, J., & McCrae, R.R. (2004). Escapable Conclusions: Toomela (2003) and the Universality of Trait Structure. *Journal of Personality and Social Psychology, 87*, 261-265. DOI: 10.1037/0022-3514.87.2.261

Austin, E.J., Deary, I.J., & Gibson, G.J. (1997). Relationship between ability and personality: Three hypotheses tested. *Intelligence, 25*, 49-70. DOI: 10.1016/S0160-2896(97)90007-6

Austin, E.J., Deary, I.J., Whiteman, M.C., Fowkes, F.G.R., Pedersen, N.L., Rabbitt, P., et al. (2002). Relationships between ability and personality: Does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences, 32*, 1391-1411. DOI: 10.1016/S0191-8869(01)00129-5

Brand, C.R. (1994). Open to experience – closed to intelligence: Why the 'Big 5' are really the 'Comprehensive 6'. *Journal of Personality, 8*, 299-310. DOI: 10.1002/per.2410080407

Corbett, J., Given, L., Gray, L., Leyland, A., McGregor, A., Marryat, L., Miller, M., & Reid, S. (2010). *The Scottish Health Survey 2008*. Edinburgh: The Scottish Government.

De Fruyt, F., Aluja, A., Garcia, L.F., Rolland, J-P., & Jung, S.C. (2006). Positive Presentation Management and Intelligence and the Personality Differentiation by Intelligence Hypothesis in Job Applicants. *International Journal of Selection and Assessment, 14*, 101-112. DOI: 10.1111/j.1468-2389.2006.00337.x

Douglas, J.W.B. (1964). *The home and the school. A study of ability and attainment in the primary school.* London: MacGibbon & Kee.

Gillespie, M., Tenvergert, E.M., & Kingma, J. (1988). Using Mokken methods to develop robust cross-national scales: American and West German attitudes to abortion. *Social Indicators Research, 20*, 181-203. DOI: 10.1007/BF00302462

Hastie, R. (1987). Information processing theory for survey researchers. In H.J. Hippler, N. Schwartz, & S. Sudman (Eds). *Social information processing and survey methodology* (pp. 42-70). New York: Springer-Verlag.

Herrman, H., Saxena, S., & Moodie, R. (eds) (2005). *Promoting mental health: concepts, emerging evidence, practice: report of the World Health Organization, Department of Mental Health and Substance Abuse in collaboration with the Victorian Health Promotion Foundation and the University of Melbourne.* Geneva: World Health Organization.

Huppert, F.A., & Whittington, J.E. (2003). Evidence for the independence of positive and negative well-being: Implications for quality of life assessment. *British Journal of Health Psychology, 8*, 107-122. DOI: 10.1348/135910703762879246

Hu, Y., Stewart-Brown, S., Twigg, L., & Weich, S. (2007). Can the 12 item General Health Questionnaire be used to measure positive mental health? *Psychological Medicine, 37*, 1005-13. DOI:10.1017/S0033291707009993

Johnson, W., Bouchard, T.J., Segal, N.L., & Samuels, J. (2005). General intelligence and reading performance in adults: is the genetic factor structure the same as for children? *Personality and Individual Differences, 38*, 1413-1428. DOI:10.1016/j.paid.2004.09.007

Karabenick, S.A., Woolley, M., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R. et al. (2007). Cognitive Processing of Self-Report Items in Educational

Research: Do They Think What We Mean? *Educational Psychologist, 42*, 139-151. DOI: 10.1080/00461520701416231

Ligtvoet, R., van der Ark, L. A., te Marvelde, J. M., & Sijtsma, K. (2010), Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, *70*, 578-595. DOI: 10.1177/0013164409355697.

Meijer, R.R., & Baneke, J.J. (2004). Analyzing psychopathology items: a case for nonparametric item response theory modelling. *Psychological Methods, 9*, 354–368. DOI: 10.1037/1082-989X.9.3.354

Messick S (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments Educational Researcher 23: 13-23. DOI: 10.3102/0013189X023002013

Molenaar, I.W., & Sijtsma, K. (2000). *MSP5 for Windows: a program for Mokken scale analysis for polytomous items* iec ProGAMMA, Groningen.

Mottus, R., Allik, J., & Pullmann, H. (2007). Does personality vary across ability levels? A study using self and other ratings. *Journal of Research in Personality, 41*, 155-170. DOI: 10.1016/j.jrp.2006.03.007,

Power, C., & Elliott, J. (2006). Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology, 35*, 34−41. DOI: 10.1093/ije/dyi183

Rammstedt, B., Goldberg, L.R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44*, 53-61. DOI: 10.1016/j.jrp.2009.10.005

Rindall, G.I., Jordhoy, M.S., Kaasa, S. (2003). Measuring quality of palliative care: psychometric properties of the FAMCARE Scale. *Quality of Life Research, 12*, 167-76. DOI: 10.1023/A:1022236430131

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. *Personality and Individual Differences, 48*, 845-48. DOI: 10.1016/j.paid.2010.02.011

Stewart-Brown, S., Tenant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-Being scale (WEMWBS): a Rasch analysis using data from the Scottich Health Education Population Survey. *Health and Quality of Life Outcomes, 7*, doi:10.1186/1477-7525-7-15. DOI:10.1186/1477-7525-7-15

Sudman, S., Bradburn, N., & Schwartz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes, 5*, 63. DOI:10.1186/1477-7525-5-63

Van der Ark, L.A. (2007). Mokken Scale Analysis in R. *Journal of Statistical Software, 20*, 1-19.

Waiyavutti, C., Johnson, W., & Deary, I.J. (2011). Do Personality Scale Items Function Differently in People with High and Low IQ? *Psychological Assessment,* DOI: 10.1037/a0026266

Watson, R. (1996). Mokken scaling procedure (MSP) applied to feeding difficulty in

    elderly people with dementia. *International Journal of Nursing Studies, 33*,

    385-393. DOI: 10.1016/0020-7489(95)00058-5

Watson, R., Deary, I.J., & Austin, E. (2007). Are personality trait items reliably more

    or less 'difficult'? Mokken scaling of the NEO-FFI. *Personality and*

    Individual Difference, 43, 1460-1469. DOI: 10.1016/j.paid.2007.04.023

Watson, R., Deary, I.J., & Shipley, B. (2008). A hierarchy of distress: Mokken scaling

    of the GHQ-30. *Psychological Medicine, 28*, 575-579. DOI:

    10.1017/S003329170800281X

Table 1 Demographic details—gender, mental ability and mental well-being—of the sample

	n	Gender (male/female)	Mental ability (SD)	Standardised mental ability (SD)	Warwick-Edinburgh score (SD)
Total	7510	3623/3887	45.95 (15.26)	0.186 (0.945)	49.25 (8.09)
Low mental ability	929	510/419	19.82 (5.18)	-1.432 (0.320)	47.86 (8.80)
Medium mental ability	5002	2395/2607	44.50 (9.17)	0.097 (0.568)	49.25 (8.07)
High mental ability	1579	718/861	65.89 (4.44)	1.421 (0.275)	50.07 (7.64)

Table 2 Mokken scaling of Warwick-Edinburgh Well-Being scale (mean item scores)

Item	1	2	3	4	5	6	7	8	9	10
1 I've been feeling optimistic about the future	$3.27^\dagger$	$3.17^\dagger$	$3.26^\dagger$	$3.36^\dagger$	$3.17^\dagger$	$3.25^\dagger$	$3.32^\dagger$	3.17	$3.27^\dagger$	$3.40^\dagger$
2 I've been feeling useful	$3.56^\dagger$	$3.44^\dagger$	$3.56^\dagger$	$3.65^\dagger$	$3.48^\dagger$	$3.53^\dagger$	3.61	$3.40^\dagger$	3.58	$3.69^\dagger$
3 I've been feeling relaxed	$3.30^\dagger$	3.24	3.30	3.33	$3.34^\dagger$	$3.35^\dagger$	$3.36^\dagger$	$3.13^\dagger$	$3.25^\dagger$	$3.31^\dagger$
4 I've been feeling interested in other people	3.54	$3.36^\dagger$	3.53	3.63	$3.20^\dagger$	3.30	3.39	3.54	3.75	3.84
5 I've had energy to spare	$2.80^\dagger$	$2.81^\dagger$	2.80	$2.81^\dagger$	$2.85^\dagger$	$2.85^\dagger$	$2.86^\dagger$	$2.75^\dagger$	$2.75^\dagger$	$2.76^\dagger$
6 I've been dealing with problems well	$3.60^\dagger$	$3.45^\dagger$	$3.61^\dagger$	$3.65^\dagger$	$3.48^\dagger$	$3.64^\dagger$	3.66	$3.46^\dagger$	$3.57^\dagger$	$3.64^\dagger$
7 I've been thinking clearly	$3.72^\dagger$	$3.57^\dagger$	$3.72^\dagger$	$3.79^\dagger$	$3.65^\dagger$	$3.77^\dagger$	$3.83^\dagger$	$3.49^\dagger$	$3.67^\dagger$	$3.76^\dagger$
8 I've been feeling good about myself	3.39	3.29	3.38	$3.45^\dagger$	3.38	3.49	3.51	$3.18^\dagger$	3.28	3.40
9 I've been feeling close to other people	$3.59^\dagger$	$3.50^\dagger$	$3.60^\dagger$	$3.60^\dagger$	3.43	$3.49^\dagger$	$3.44^\dagger$	$3.58^\dagger$	$3.70^\dagger$	$3.73^\dagger$
10 I've been feeling confident	3.46	3.33	3.46	3.55	3.45	3.56	$3.62^\dagger$	$3.18^\dagger$	3.37	3.49
11 I've been able to make up my own mind about things	$3.96^\dagger$	$3.91^\dagger$	$3.96^\dagger$	$4.00^\dagger$	$3.94^\dagger$	$4.01^\dagger$	$4.04^\dagger$	$3.87^\dagger$	$3.92^\dagger$	$3.97^\dagger$
12 I've been feeling loved	$3.91^\dagger$	$3.57^\dagger$	$3.91^\dagger$	$4.00^\dagger$	$3.76^\dagger$	$3.84^\dagger$	$3.80^\dagger$	$3.89^\dagger$	$3.99^\dagger$	$4.04^\dagger$
13 I've been interested in new things	$3.60^\dagger$	3.44	$3.59^\dagger$	$3.93^\dagger$	$3.49^\dagger$	$3.60^\dagger$	$3.68^\dagger$	$3.38^\dagger$	$3.57^\dagger$	$3.75^\dagger$
14 I've been feeling cheerful	3.58	3.51	3.58	3.60	3.54	3.58	$3.56^\dagger$	$3.47^\dagger$	$3.58^\dagger$	$3.63^\dagger$
H	0.48	0.44	0.49	0.49	0.44	0.48	0.48	0.46	0.52	0.52
H^T	0.31	0.23	0.32	0.35	0.22	0.30	0.34	0.25	0.39	0.43
Rho	0.91	0.90	0.92	0.91	0.90	0.91	0.91	0.91	0.92	0.92
No. of items in scale	10	9	8	10	10	10	10	12	10	11

1 Whole sample; n=7510; p=0.0003
2 Low mental ability; n=929; p=0.0003
3 Medium mental ability; n=5002; p=0.0003
4 High mental ability; n=1579; p=0.0003
5 Males, low mental ability; n=510; p=0.0003
6 Males, medium mental ability; n=2395; p=0.0003
7 Males, high mental ability; n=718; p=0.0003
8 Females, low mental ability; n=419; p=0.0003
9 Females, medium mental ability; n=2607; p=0.0003
10 Females, high mental ability; n=861; p=0.0003

†=items showing invariant item ordering; Note: H= Loevinger's coefficient, a measure of the ratio of the observed to expected errors in the order or scalability of items which ranges from 0 (no scalability) to 1 (perfect scalability). H^T is a measure of the ratio of observed to expected violations of IIO. Rho is analogous to test-retest reliability