promoting access to White Rose research papers

White Rose
Research Online

Universities of Leeds, Sheffield and York
http://eprints.whiterose.ac.uk/

Short communication

Eysenck's EPI, Goldberg's IPIP and the hierarchy of items within personality trait scales

Summary

Based on the recent finding of a hierarchical scale for Neuroticism in the NEO-Five Factor Inventory, two further personality inventories: the Eysenck Personality Inventory and Goldberg's International Personality Item Pool were analysed using the Mokken Scaling Procedure for hierarchical scales. Items from two dimensions of the Eysenck Personality Inventory: Neuroticism and Extraversion produced hierarchical scales of 12 and five items, respectively. The Neuroticism items ran from items expressing mild to more extreme worry and the Extraversion items ran from mild sociability to more extreme 'showing off'. The utility of hierarchical scales in personality measurement is discussed in terms of furthering theoretical understanding of personality and also practical application. In addition, the reasons why only one of these scales should produce hierarchical sets of items is discussed.

1. Introduction

A recent paper in *Personality and Individual Differences* (Watson, Deary, & Austin, 2007) demonstrated that some items within personality trait scales form a hierarchy. Specifically, items from the Neuroticism scale of the NEO-Five Factor Inventory (NEO-FFI; Costa & McRae, 1992) formed a hierarchy whereby participants pooled from several large UK studies (n=1028) more readily endorsed that they were 'stressed' and 'going to pieces' than that they were 'hopeless' and wanted someone to 'solve problems' for them, with other items intervening between these. In the terminology of hierarchical scales, items which are less readily endorsed are referred to as having greater 'difficulty' (Watson, Deary, & Austin, 2007) and, thereby, represent a greater amount of the latent trait being measured. In the case of the Neuroticism scale of the NEO-FFI, greater difficulty means higher levels of Neuroticism.

Hierarchies among scale items represents a different - complementary and useful - way of looking at the relationship between item scores and scale total scores, which is the purpose of methods emanating from item response theory. In this way, individuals' scores on a questionnaire - 'the observable data' (Hulin *et al.* 1983, p. 14) - can be related to the underlying theoretical construct that is being measured. The method of choice for investigating the possibility of hierarchical scales in multivariate data is Mokken scaling (Mokken & Lewis, 1982). The application of this procedure has been greatly facilitated by the development of a computer programme for the analysis: the Mokken Scaling Procedure (MSP; Sijtsma *et al.* 1990). The details of Mokken scaling and its application using the MSP to data from the NEO-FFI are explained by Watson, Deary, and Austin (2007). Briefly, the MSP provides a series of

diagnostics which allows the investigator to establish whether reliable hierarchies of items exist in a multivariate data set and, thereafter, to check that these items have both monotone homogeneity and are doubly monotonous. This establishes that the score on an item increases as the score on the latent trait increases and that the item-response curve for the items that are included in the hierarchical scale do not intersect.

Given the identification of a hierarchical scale for Neuroticism in the NEO-FFI it is important to discover whether or not such hierarchies exist in other personality inventories. The present study investigates two such inventories: the Eysenck Personality Inventory (EPI; Eysenck & Eysenck 1964) and Goldberg's IPIP (International Personality Item Pool) Big-Five factor markers (Goldberg 1999).

2. Methods

2.1 Participants

2.1.1 EPI

The 9003 (3905 men, 5098 women) participants completing the Eysenck Personality Inventory (EPI) came from the Health and Lifestyle Survey (HALS) which is a UK nationwide sample survey of community-dwelling adults resident in England, Scotland, and Wales. Mean age of the sample was 45.9 years (sd=17.7). Full details of the study can be seen elsewhere (Cox, Blaxter, Buckle, Fenner, Golding, Gore, Huppert, Nickson, Roth, Stark, Wadsworth & Whichelow, 1987; Shipley, Weiss Der, Taylor & Deary, 2007). The EPI (Eysenck & Eysenck, 1964) is a self-report personality inventory consisting of 57 items measuring Extraversion and Neuroticism. The response format of the EPI's items involves marking 'yes' or 'no' to each statement. Nine of the 57 questions formed a lie scale. Scores range from 0 to 24 for

each personality trait with higher scores representing higher Neuroticism or Extraversion. The questionnaire was completed at home and returned by post. Test-retest reliabilities of the scale based on normal samples are excellent at 0.84 for Neuroticism and 0.88 for Extraversion with a time lapse of one year between test and retest (Eysenck & Eysenck 1964).

2.1.2 IPIP

Participants completing the IPIP came from the Lothian Birth Cohort 1921 (LBC1921), whose initial recruitment and testing is described elsewhere (Deary, Whiteman, Starr, Whalley, & Fox, 2004). All were born in 1921 and are surviving participants of the Scottish Mental Survey of 1932. They were recruited into the LBC1921 (a longitudinal study of cognitive ageing) at a mean age of 79.1 years (sd = 0.6). The subsequent distribution of the IPIP questionnaire to the LBC1921 has been described previously (Gow, Whiteman, Pattie, & Deary, 2005). As part of ongoing follow-up, they were mailed the 50-item version of the IPIP Big-Five factor markers to assess personality when aged around 81 years old. IPIP questionnaires were sent to 534 participants, of which 498 were returned (93%). The sample sizes with complete data for each factor were Extraversion: male = 192, female = 268; Agreeableness: male = 192, female = 273; Intellect: male = 192, female = 270; Conscientiousness: male = 193, female = 269; Emotional Stability: male = 190, female = 271.

The IPIP scale contains 10 items for each of the Big-Five personality factors: Extraversion (E), Agreeableness (A), Conscientiousness (C), Emotional Stability (ES) and Intellect (I). Participants were instructed to read the 50 items and mark each one according to how much they believed it described them on a five point scale (very

inaccurate to very accurate). The IPIP scales correlated moderately to highly with the appropriate scales from another short-form five-factor inventory, the NEO-FFI (Costa & McCrae, 1992): the IPIP-NEO scale correlations are 0.69 (E), 0.49 (A), 0.76 (C), -.83 (ES/Neuroticism) and 0.59 (I/Openness; Gow, Whiteman, Pattie & Deary, 2005). Cronbach's alphas in this sample are 0.84 (E), 0.76 (A), 0.77 (C), 0.87 (ES) and 0.73 (Gow, Whiteman, Pattie & Deary, 2005).

2.2 Statistical analysis

Both sets of data from participants who completed the EPI and the IPIP were entered into an SPSS version 13.0 database. These data were saved in a tab-delimited form with the 'write variable names to spreadsheet' option deselected. Any individuals with missing data were then removed from the database to prepare the data for analysis using the MSP, leaving 5795 participants who completed the EPI, and 450 who completed the IPIP. Mokken scaling was carried out on resulting data sets (all participants and all items) using the MSP version 5.0 for Windows (Molenaar & Sijtsma 2000). The procedure for running the MSP and selecting items has already been described by Watson, Deary, & Austin (2007).

3. Results

Items from the IPIP did not produce any reliable Mokken scales. On the other hand, the EPI produced two Mokken scales as shown in Table 1. One scale of EPI items was composed entirely of 12 items related to Neuroticism. The other scale of EPI items was composed entirely of five items related to Extraversion. Both scales demonstrated acceptable Mokken scaling parameters of scalability (H > 0.4), reliability (Rho ≈ 0.7) and probability (p < 0.05). Items in the Neuroticism scale run

from mild expressions of Neuroticism such as worrying and being easily hurt at the lower end of the scale to having attacks of shaking and having nightmares at the higher end of the scale. Items in the Extraversion scale run from liveliness at the low end of the scale to being the 'life and soul of a party' and 'doing anything for a dare' at the higher end.

4. Discussion

Mokken scaling has again proved to be a useful analytical technique for demonstrating that personality traits, at least in the case of the EPI, can be measured hierarchically. These novel findings complement the more usual factor-analytic approaches to the traits. It is also interesting to note, in the case of the EPI, that items from two personality dimensions form hierarchies and that there was no overlap of items between dimensions in the hierarchical scales. It was already known that the developers of the EPI produced robust scales in which the relationship between items and latent traits is distinct. It appears that, as with the NEO-FFI (Watson, Deary, & Austin 2007), they have also, and perhaps inadvertently, produced scales in which the items related to these latent traits are hierarchical.

In addition to providing measures of the extent to which a latent trait is present, the added value of Mokken scaling is that the presence of particular aspects of personality traits also indicates what other aspects of the trait are likely to be present. For example, with reference to Neuroticism, if someone is plagued by feelings of guilt then they are also likely to feel inferior and to have their feelings easily hurt.

Why should some scales demonstrate hierarchies and others not? Clearly, it must be related to the content of the scales, to the items that compose the scales and the extent to which they enable measurement of different levels of the purported latent traits. Some aspects that might contribute to the existence of a Mokken scale are: having many items; having items that tap a similar, narrow dimension; and having items with a large spread of difficulty or endorsements. The EPI contains, within each dimension, many more items than the IPIP. Both the EPI and IPIP Neuroticism scales, for example, contain items related to anxiety and depression. The EPI has additional items on physical symptoms (for example, shaking, palpitations), some of which are at the more severe end of the Neuroticism dimension. Examining the mean item responses of the EPI and IPIP, and comparing these with the item standard deviations indicates a tendency for EPI items to have more spread in their endorsement levels. Therefore, it might be the case that the EPI's greater number of items and greater spread of severity among its items accounts for many of its items, especially in the Neuroticism scale, having a hierarchical nature. The reliability and validity of the IPIP is not being questioned. We simply note that the EPI shows some additional characteristics in having hierarchical scales.

Mokken scaling adds theoretical information and possibly practical utility to personality scales. It is theoretically valuable because it finds a reliable hierarchy of phenomenology (mostly self-reported) on trait dimensions. It tells us more about what the trait 'feels' like at different points on the scale. It could be practically valuable too. Adaptive testing is useful in cognitive assessment. If more work were done to find robust hierarchies of items on personality scales, then adaptive testing of

personality traits could be possible also.

References

Costa, P.T., & McRae, R.R. (1992) *Revised NEO personality inventory and NEO-five-factor inventory: Professional manual* Psychological Assessment Resources, Inc, Odessa FL USA.

Cox BD, Blaxter M, Buckle ALJ, Fenner NP, Golding JF, Gore M, Huppert FA, Nickson J, Roth M, Stark J, Wadsworth MEJ, Whichelow M (1987)*The Health and Lifestyle Survey: A Preliminary Report.* Health Promotion Trust, London.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 130-147.

Eysenck SB, Eysenck HJ. (1964) An improved short questionnaire for the measurement of extraversion and neuroticism. *Life Sciences*, 305,1103-1109.

Goldberg, L.R. (1999) A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models. In I. Mervielde, I.J. Deary, F. de Fruyt, & F. Ostendorf (Eds.). *Personality psychology in Europe* (Vol. 7, pp 7-28). Tilburg: Tilburg University Press.

Gow, A.J., Whiteman, M.C., Pattie, A., & Deary, I.J. (2005) Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39, 317-329.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983) Introduction to item response theory in Hulin, C.L., Drasgow, F., & Parsons, C.K *Item response theory* Dow Jones-Irwin, Homewood, Illinois, pp 13-74.

Mokken, R.J., & Lewis, C. (1982) A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement 6, 417-430.*

Molenaar I.W., & Sijtsma K. (2000). *MSP5 for Windows.* Groningen: iec ProGAMMA.

Shipley, B.A., Weiss, A., Der, G., Taylor, M.D., & Deary, I.J. (2007) Neuroticism, extraversion and mortality in the UK Health and Lifestyle Survey: 21 year prospective cohort study. Psychosomatic Medicine, 69, 923-931.

Sijtsma, K., Debets, P., & Molenaar, I.W. (1990) Mokken scaling analysis for polychotomous items: theory, a computer programme and an empirical application. *Quality & Quantity 24, 171-188.*

Watson, R., Deary, I., & Austin, E. (2007) Are personality trait items reliably more or less 'difficult'? Mokken scaling of the NEO-FFI *Personality and Individual Differences* **43** 1460-1469

Table 1

Mokken scale of EPI checked for violations of monotone homogeneity and double montonicity

EPI Item (or short paraphrase of item)	Mean	H
Neuroticism scale[†]		
43. have nightmares	1.11	0.33
35. get attacks of shaking	1.12	0.38
47. nervous person	1.23	0.42
26. think you are tense	1.24	0.41
23. often troubled by guilt	1.29	0.37
52. feel inferior	1.30	0.38
2. need friend to cheer you up	1.39	0.37
40. worry about awful things that may happen	1.45	0.38
9. feel 'just miserable' sometimes	1.45	0.36
16. feelings easily hurt	1.60	0.45
50. easily hurt when people fault your work	1.60	0.42
14. worry about things shouldn't have done	1.63	0.46
Extraversion scale[‡]		
10. do anything for a dare	1.11	0.44
53. get some life into dull party	1.35	0.57
29. quiet when with other people*	1.52	0.41
27. other people think lively	1.56	0.49
51. hard to enjoy lively party*	1.70	0.53

* these items are reverse scored

† Scale H = 0.40; p < 0.001; Rho = 0.82

‡ Scale H = 0.49; p < 0.001; Rho = 0.69