British Journal of Preventive and Social Medicine, 1977, **31**, 122-126

Scaling indices of disablement

A. C. BEBBINGTON

From the Personal Social Services Research Unit, University of Kent at Canterbury

SUMMARY Williams *et al.* (1976) have suggested the use of Guttman scaling for scoring an index of disability. Two examples confirm the applicability of this method in the context of survey research. One of these examples is of a disablement scale widely employed in local authority social services research. For the purpose of survey assessment of disabled populations, the precise choice of scaling method for scoring disability is often of little consequence.

Williams *et al.* (1976) have recently again raised the issue of what numerical values should be assigned as weights in the construction of an additive index of self care disablement. They have proposed a model of cumulative disablement to which Guttman scale analysis can be applied.

A particular case which they mention is an assessment of personal dependency used in a national survey of the disabled in England and Wales (Harris, 1971). The principal survey question on which this assessment was based was subsequently widely employed in local authority surveys after the 1971 Chronically Sick and Disabled Persons (CSDP) Act. It is given in Table 1.

It is the contention of the current paper that, for the purpose of obtaining a broad picture of the disabled population, the choice of numerical values is, within reason, irrelevant for a scale of this type. That is to say, any reasonable set of weights can be chosen arbitrarily without affecting greatly the properties of the resulting index, so that, for example, discussion on whether being able to feed oneself is twice as important to disablement as being able to dress oneself is unnecessary for this purpose. This will be true for any scale that is inherently unidimensional in the sense in which it is discussed below, and this paper shows that this is so for the question in Table 1 based on a survey of Kensington and Chelsea for the CSDP Act 1971 (Buckle and Baldwin, 1972). A scoring system using an additive index is compared with Guttman scaling and principal components scaling.

The results obtained from this question and those from another disablement scale confirm, with larger sample sizes, the findings of Williams *et al.* (1976) that Guttman scaling is a suitable method for scales of this type.

Table 1 *Personal dependency question (Harris, 1971), showing the item weights for the normative scale as used in the Kensington and Chelsea survey*

Do you generally have difficulty	*No difficulty or supervision*	*If difficulty or supervision ask: Can you do it yourself, even with difficulty?*	
		Yes can do	*No cannot do*
(i) Getting in and out of bed on your own?	0	2	3
(ii) Getting to or using the WC?	0	4	6
(iii) Having an all over wash (or bathing yourself if bath used)?	0	2	3
(iv) Washing your hands and face?	0	2	3
(v) Putting on shoes and socks or stockings yourself?	0	2	3
(vi) Doing up buttons and zips yourself?	0	4	6
(vii) Dressing, other than buttons and shoes?	0	2	3
(viii) Feeding yourself?	0	4	6
(ix) Women and children only: Combing and brushing your hair? Men only: Shaving yourself?	0	2	3

The normative scale

The method of assessing the personal dependence of survey subjects in the Kensington and Chelsea survey is by an additive index using the weights given in Table 1. Similar methods were widely employed in many CSDP Act surveys. This method 'corresponds exactly with criteria used by social workers within the borough for defining handicap' (Buckle and Baldwin, 1972) and for this reason we refer to it as the normative scale. It relies on the classification of the items into major or minor disability; the former (such as being able to feed oneself) is given twice the weight of the latter (such as being able to dress oneself). Each item is allocated a score on a scale, 0, 2, 3 or 0, 4, 6 which indicates the value judgement that the difference between being able to do something and having difficulty with it is more important than the difference between having difficulty and not being able to do it at all.

So there are two arbitrary elements in the normative scale: the relative importance of each item in the weighting system, and the scoring system of categories within each item.

The class of 'reasonable' alternatives to the normative scale may be any scale which is an additive combination in which the categories within items have weights in ascending sequence.

Guttman scaling

Most textbooks on scaling give an account of the Guttman method (see, in particular, Stouffer *et al.*, 1950). The basic principle in the present context is to determine whether a set of items can be put in a hierarchy of severity. Can items be graded from easy to hard, in such a way that any subject who can perform a particular task will certainly be able to perform all tasks rated easier, and, conversely, if he cannot perform a particular task, he will certainly be unable to perform any task rated as harder?

For reasons of computational ease, a slightly simplified form of Guttman scaling has been used to scale the nine, three-category items, compared with, say, the method recommended by Stouffer *et al.* (1950) (chapter 4) for analysing multi-category items. We have turned the nine items into 18 two-category items. Each item is turned into two, in the following way:

(i) Is this task done without difficulty or supervision? (yes/no)

(ii) Can this task be done at all (even with difficulty)? (yes/no)

On this basis an order to these items can be found which satisfies the Guttman scaling criterion. This is given in Table 2. A perfect Guttman scale, as described above, does not exactly obtain, but the number of 'errors' is small. The Guttman criterion of reproducibility is 0·94, calculated over 18 items, and correspondingly the coefficient of scalability (Menzel, 1953) is 0·53. The recommended weights for the scale are identical by either method. They are 0, 1, 2 for each category respectively, being the same for every item. The scale is formed as an additive index.

Table 2 *Order of items in the personal dependency question given by Guttman scaling, and the severity grading scale*

Item	*Percentage of sample for whom this is true*	*Severity grading scale*
No difficulty with any task	55	0
Has difficulty		
with having an all over wash	33	1
putting on shoes and socks	27	2
doing up buttons and zips	21	3
getting in and out of bed	20	4
dressing	18	5
getting to or using WC	17	6
brushing hair/shaving	15	7
washing hands and face	10	9
feeding self	8	11
Cannot manage		
having an all over wash without help	14	8
putting on shoes and socks without help	9	10
getting to or using WC without help	8	12
doing up buttons and zips without help	7	13
dressing without help	6	14·5
getting in and out of bed without help	6	
brushing hair/shaving without help	4	16
washing hands and face without help	3	17
feeding without help	1	18

Sample size = 377
Note that 'has difficulty with' includes 'cannot manage'.

Reproducibility is calculated by the Goodenough method, as used in the SPSS programme (Anderson, 1966) for dichotomous variables which are known to give slightly lower estimates of reproducibility than do other methods on dichotomous variables. The adaption for trichotomous variables given here should result in estimates of reproducibility not dissimilar from those given by other multi-category methods of Guttman scaling.

Severity scale

A minor variant of the conventional Guttman scale is to score each person, not by the addition of weights, but according to the 'easiest' task they are unable to do. This scale is given in the right-hand column of Table 2. For example, a person who cannot dress himself, even with difficulty, but can do all the subsequent easier tasks in Table 3 such

as feeding himself, would be given a severity score of 13.

Since the Guttman scale has identified a clear order for these items, Williams *et al.* (1976) equate this order with progressive stages in a sequence of physical decline, and we refer to this as a severity scale. It is not too dissimilar from the ranking of physical state produced by Wright (1974) who argued the virtues of a ranking scheme in contrast to an additive index.

Principal component scaling

Like the Guttman scaling, principal component scaling is a method that relies on the internal evidence of the data being scaled, rather than on normative judgement, in order to produce a scale. Again the reader is referred to a textbook such as Morrison, 1976 (chapter 8) for full details. The method calculates the relative weights to be given to each item to construct a linear additive index which contains the maximum information (in the technical sense of the greatest amount of variance) of all possible such indices. The question is whether a set of weights can be found for those items which will give a single scale containing virtually all the information in the nine items separately and which can be used to account for most of the variation between individuals?

There is the slight complication that principal components scaling demonstrates a relative pattern of weights between items, but does not scale the categories of each item. There are variations on this method adapted for this purpose (Healy and Goldstein, 1976). For simplicity, we have left the categories scaled as for the normative scale which was 0, 2, 3. An additional analysis compared the result of using ratios 0, 2, 3 with the results of the two extreme schemes: 0, 0, 1 and 0, 1, 1 and found only moderate differences for these data.

Principal components scaling produced a scale that accounts for 62% of the total variance. This weighting scheme is derived from the first principal component of the correlation matrix of items. (Incidentally, none of the other eight principal components accounts for more than 8% of the total variance.) This reasonably satisfies the requirement for reducing these items to a single principal component scale. The weights are given in Table 3.

Choice of scale

It will now be demonstrated that in the present case the choice of scaling method, despite the variation in the weights and hence in the relative importance given to different items, is irrelevant in terms of the final scale produced. This is done quite simply by considering the correlations between scale scores produced by the different methods, given in Table 4. The product moment correlations between these scales are in every case greater than 0·90 indicating they are all measuring the same thing. The only respect in which they differ is

Table 4 *Product moment correlations between the four scales produced by different methods of personal dependency question (Harris, 1971)*

Scale	*(i)*	*(ii)*	*(iii)*	*(iv)*
(i) Normative	1·00			
(ii) Guttman	0·99	1·00		
(iii) Severity	0·91	0·90	1·00	
(iv) Principal components	0·99	0·99	0·90	1·00

Table 3 *Personal dependency question (Harris, 1971), showing the item weights generated by the version of principal components scaling used*

Do you generally have difficulty	*No difficulty or supervision*	*If difficulty or supervision ask: Can you do it yourself, even with difficulty?*	
		Yes can do	*No cannot do*
	(1)	(2)	(3)
(i) Getting in and out of bed on your own?	0	1·61	2·41
(ii) Getting to or using the WC?	0	1·67	2·51
(iii) Having an all over wash (or bathing yourself if bath used)?	0	1·15	1·73
(iv) Washing your hands and face?	0	2·02	3·04
(v) Putting on shoes and socks or stockings yourself?	0	1·48	2·22
(vi) Doing up buttons and zips yourself?	0	1·61	2·42
i(vii) Dressing. other than buttons and shoes?	0	1·95	2·93
(viii) Feeding yourself?	0	2·52	3·78
(ix) Women and children only: Combing and brushing your hair? Men only: Shaving yourself?	0	1·96	2·94

with regard to their means and standard deviations, which can in any case be arbitrarily chosen.

Table 5 gives the mean score on each of the four indices, after they have been standardised to a common mean and standard deviation, for three age groups. The differences are slight and it is clear that the inference about the relationship between age and personal dependency would have been the same whichever scaling method was used. The apparent decline of personal dependency with age is an artefact due to the way the survey population is defined. The 'survey population' for the personal dependency questions consisted of all people in the borough of Kensington and Chelsea who described themselves as impaired or handicapped (including blindness, deafness, immobility, etc.), or who were over 75 years and living alone. However, people under 70 with a low degree of handicap were subsequently excluded. For a full account, see Buckle and Baldwin (1972).

Table 5 *Mean score by age group for four scales produced by different methods from personal dependency question (Harris, 1971)*

Age group	*Mean score on standardised scale*		
	Up to 64	*65-74*	*Above 75*
(i) Normative scale	0·439	0·067	−0·182
(ii) Guttman scale	0·415	0·062	−0·172
(iii) Severity scale	0·442	0·118	−0·199
(iv) Principal components scale	0·425	0·087	−0·184
Sample size	(83)	(69)	(225)

For comparison, each scale has been standardised.

Unidimensionality

The conclusion that choice of scaling is often irrelevant is expected to hold good for any scale which consists of a number of similar items, each with a small number of categories which have a clear-cut ordering, provided the condition for principal components scaling obtains—that is, the first principal component accounts for a very substantial part of the total variance, while the remaining components account for small, roughly equal, parts of the total variance. This is the criterion for unidimensionality.

Any scale which is suitable for Guttman scaling, that is one which has a high coefficient of reproducibility, is also likely to be unidimensional. An intuitive explanation for this is as follows. There is a very close relationship between the amount of variance accounted for by the first principal component and the average product-movement correlation between items in an index, assuming these correlations are generally positive. The higher the average correlation, the greater the importance of the first principal component, and hence the greater likelihood of unidimensionality. Consider two, two-category items. If these scale between themselves with perfect reproducibility then one cell of their four-celled cross tabulation table will be vacant. But this is precisely the condition under which the product moment correlation between two, two-category items is maximised, subject to fixed marginals. Hence for a scale of several two-category items, perfect reproducibility implies that the first principal component is maximised, subject to the marginal distributions of the items, and so unidimensionality is likely to obtain.

If Guttman scaling proves successful for a scale of this type, it is probably unnecessary to consider more sophisticated weighting systems. However, a word of caution needs to be introduced concerning inferences based on the size of the coefficient of reproducibility, since this coefficient depends not only on the underlying scalability of the index but also on the marginal distribution of the items. Neither does this coefficient on its own give any information on the error structure, whether, for example, errors occur at random; nor on whether certain items are unsuitable. For indices of disability the marginal distribution of items is particularly likely to be extreme, and this will artificially raise reproducibility (Stouffer *et al.*, 1950, chapter 3). An independent check of the unidimensionality of the scale, even by a rough-and-ready method, is recommended.

A further application of Guttman scaling

We confirm the general applicability of Guttman scaling to disability scales with one further example. Table 6 illustrates the results of applying the method to an index of domestic self care ability applied to a general sample of 845 old people in Glamorgan (Watson and Albrow, 1973). The

Table 6 *Order of items from Guttman scale of domestic dependency from a survey by Watson and Albrow*

Item	*Percentage of respondents for whom this was true*
No difficulty with any task	30
Difficulty with, or unable to manage	
general repairs	62
window cleaning	52
heavy cleaning	46
doing the laundry	37
shopping	34
dusting and tidying	29
getting meals	26
Sample size (no.)	(845)

The question was: 'Would you tell me if you find difficulty with, or are unable to manage, any of the following chores, and whether anyone helps you, or does them for you?'

coefficient of reproducibility here was 0·93 and a principal component accounting for 63% of the variance (with no other significant components) indicated satisfactory unidimensionality. Menzel's coefficient of scalability was 0·81.

I should like to thank Kensington and Chelsea Social Services Department, and Margaret Watson of University College, Cardiff, for the use of data reported in this paper, and Rose Knight of the University of Kent, for advice in the formulation of these ideas.

Reprints from A. C. Bebbington, Personal Social Services Research Unit, University of Kent, Canterbury.

References

Anderson, R. E. (1966). A computer program for Guttman scaling with the Goodenough technique. *Behavioural Science*, **11**, 235.

Buckle, J., and Baldwin, P. (1972). *Survey of Handicapped and Impaired Persons and Persons Aged 75 or Over and Living Alone*. Royal Borough of Kensington and Chelsea Social Services Department: London.

Harris, A. I. (1971). *Handicapped and Impaired in Great Britain*. Office of Population Censuses and Surveys: London.

Healy, M. J. R., and Goldstein, H. (1976). An approach to the scaling of categorised attributes. *Biometrika*, **63**, 219-229.

Menzel, H. (1953). A new coefficient for scalogram analysis. *Public Opinion Quarterly*, **17**, 268-280.

Morrison, D. F. (1976). *Multivariate Statistical Methods*, 2nd edition. McGraw Hill: New York.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., and Clauson, J. A. (1950). *Measurement and Prediction*. Wiley: New York.

Watson, M., and Albrow, M. (1973). *The Needs of Old People in Glamorgan*. Glamorgan County Council Social Services Department: Cardiff.

Williams, R. G. A., Johnston, M., Willis, L. A., and Bennett, A. E. (1976). Disability: A model and a measurement technique. *British Journal of Preventive and Social Medicine*, **30**, 71-78.

Wright, K. G. (1974). Alternative measures of the output of social programmes: The elderly. In *Economic Policies and Social Goals: Aspects of Public Choice*. Edited by A. J. Culyer. Martin Robertson: London.