Social Indicators Research (2007) 82: 433–442
DOI 10.1007/s11205-006-9040-z

W. C. IP, Y. K. KWAN and L. L. CHIU

MODIFICATION AND SIMPLIFICATION OF THURSTONE SCALING METHOD, AND ITS DEMONSTRATION WITH A CRIME SERIOUSNESS ASSESSMENT

(Accepted 16 June 2006)

ABSTRACT. The paper aims to improve the Thurstone scaling method by reducing the workload of data collection and simplifying the procedures of application. It proposes a hierarchical structure for organizing items that are to be scaled. Instead of making paired comparisons, the respondents would be asked to rank the items. The ranks of the items would then be transformed into paired comparisons. Workload of data collection can hence be greatly reduced. As an alternative to Thurstone model, the use of three models, i.e., linear, exponential and information, to compute serious scores is also demonstrated.

KEY WORDS: continuum scale, hierarchical structure, paired-comparison, seriousness ranking

1. INTRODUCTION

Thurstone (Thurstone, 1927, 1928; Thurstone and Chave, 1929) scaling method of attitude has been widely used in social sciences, particularly in measuring perceived seriousness of health symptoms (Hunt et al, 1986; Bowling, 1997; Ip et al, 1999) and perceived crime severity (Kwan et al, 2000, 2002). The method, though theoretically sophisticated and vigorous and practically successful and popular, has been attacked on two fronts: it is very demanding on data and its method is complicated and tedious to apply. The method uses paired comparisons, and hence, to assess the relative importance of, say, 15 selected items, 100 replications will be needed and this will involve 10,500 paired comparisons. Unlike other elementary statistical techniques that can be confidently handled by most social scientists, some special training in statistics and computing techniques is needed in order to apply the method.

The paper aims to simplify the Thurstone method (referred as 'Original Method' hereafter) by reducing the workload of data collection and

simplifying the procedures of application. It will first show that by organizing items into a hierarchical structure and by collecting ranking data and transforming them into paired comparisons (referred as 'Simplified Method' hereafter), workload of data collection can be greatly reduced. An experiment will then be performed to demonstrate the validity of the Simplified Method in crime severity evaluation. It also attempts to show that by incorporating Thurstone model into Latent Models (Lazarsfeld, 1950; Bartholomew, 1993) and taking it as a branch, some simple models that can easily be handled even by non-specialists in statistics can be formulated.

2. REDUCING THE WORKLOAD OF DATA COLLECTION

Scaling methods are often used in measuring Quality-of-life (QOL). Unfortunately QOL instruments usually involve many health items. For examples, the Sickness Impact Profile (SIP) has 136 items (Bergner et al., 1992), the Nottingham Health Profile (NHP) has 45 symptoms (Hunt et al., 1986), the Short-form-36 (SF-36) has 36 items (Ware and Sherbourne, 1992; Medical Outcomes Trust, 1993) and the HSI has 28 items (Ip et al., 1999). To calculate the proportions of the health items that are judged to be more serious than others by 100 replications, the SIP would need 918,000 paired comparisons, the NHP, 99,000 paired comparisons, the SF-36, 63,000 paired comparisons, and finally the HSI, 37,800. Fortunately the health items can be classified into different domains, and this reduces very much the number of paired comparisons required. Just for convenience, assuming 11 items for each of the 12 domains of the SIP, roughly speaking, the number of paired comparisons would have been reduced to 66,000. However, it is still a very large number. Moreover in doing so, there can only be a total score for each of the domains while it is not possible to calculate an aggregate score that represents all the 12 domains. This is the reason why an overall score is missing for all the SIP, NHP and SF-36 QOL instruments.

To reduce the heavy burden of data collection, the Simplified Method with two modifications may be introduced. First, instead of asking a respondent to compare the relative seriousness of two health items at each time, a respondent can be required to rank the items in an area. Suppose there are, say, 6 items, A, B, C, D, E and F in an area. Originally there are 15 paired comparisons for each replication. Now a respondent is requested to rank the 6 items. The ranking can then be broken down into 15 paired comparisons. For example, a respondent gives the ranking, $A > B > C > D > E > F$ ($>$ means more serious). The ranking can be reduced to the following 15 paired comparisons: $A > B$, $A > C$, $A > D$, $A > E$, $A > F$, $B > C$, $B > D$,

B > E, B > F, C > D, C > E, C > F, D > E, D > F, and E > F. Moreover, ranking can avoid inconsistency that may occur in paired comparisons. If a respondent judged A as more serious than B, B more serious than C, and C as more serious than A, then the respondent contradicts himself/herself in the three judgements. This kind of inconsistency can be avoided by using ranking. However, it is easier for a respondent to make pair-wise comparison than to rank a number of items, and the difficulty in ranking increases with the number of items. Hence the number of health items within a domain cannot be too large and should be kept as small as possible.

To avoid too many items to be ranked we propose the second modification: a hierarchical-structural design for the health items. Taking SIP as an example, the 136 health items may be regarded as the bottom level. The 12 domains that contain the health items may be regarded as the second level. The domains in the second level may further be grouped into, say, areas in the third level, and the hierarchical structure can be represented by a tree diagram. Respondents are requested to rank the areas in the top level, the domains within each area in the second level, and finally the health items within each domain in the bottom level. From the structural rankings weights for items in each level can be obtained. Multiplying the weights from the top level through to the bottom would then give the relative importance of each health item. An application of such structure may be found in Ip et al. (1999).

3. ORGANIZING HEALTH ITEMS INTO A HIERARCHICAL STRUCTURE

The HSI instrument in Ip et al. (1999) consists of seven sections, each having four symptoms, making a total of 28 symptoms. To utilize Thurstone scaling method for paired comparisons in deriving weights of seriousness of perceived symptoms, each respondent will be required to make $\binom{28}{2} = 238$ paired comparisons among these symptoms. This is definitely too much demanding for a respondent. To reduce the number of paired comparisons that a respondent need to make and streamline the interview Ip et al. (1999) have adopted a hierarchical structure.

At the uppermost level, the symptoms are divided into two broad areas, namely, 'Physical' and 'Socio-psychological'. Respondents are asked to rank the seriousness of these two broad areas according to their seriousness in affecting one's health status. Within the 'Socio-psychological' area, the respondents are asked to rank the seriousness of the sections 'Mood' and 'Social link'. Likewise, the 'Physical' area is subdivided into three sections, in

which the 'Physical decline' section is further subdivided into yet three subsections. In such a way, a respondent need only to rank the seriousness of a few sections at the same division level at a time. At the bottom level, a respondent will be asked to rank the seriousness of the four symptoms in each of the seven sections. Hierarchical structure is a powerful tool to reduce data demand. Even without using the ranking technique, the total number of paired comparisons that a respondent has to make can be reduced to $7 \times \binom{4}{2} + 2 \times \binom{3}{2} + 2 = 50$.

4. DEVELOPING SCALING METHODS

Suppose the perceived relative seriousness of a set of n items is to be measured. Let S be the unknown underlying scale of seriousness, and s_i be the scale value for item i. Define (x_i, x_j) as the outcome of a paired comparison between items i and j:

$$(x_i, x_j) = \begin{cases} 1, & \text{if } i \text{ is judged more serious than } j \\ 0, & \text{otherwise} \end{cases}.$$

Let $\pi(s_i, s_j)$ be the probability that $(x_i, x_j) = 1$ for some given scale values s_i and s_j. It can be readily shown that the sufficient and necessary condition for $s_i > s_j$, is for any i,j $\sum_{k=1}^{n} (x_i, x_k) > \sum_{k=1}^{n} (x_j, x_k)$. Ip et al. (2004) have shown that the results from pair-wise comparisons will also provide a seriousness ranking of the items and this ranking scale will give an ordinal measurement for the item severity.

They have also suggested several models to transform the seriousness probability to a seriousness score. All of them offer some flexibility in modeling the underlying relation between the observable judgments and the underlying attitude scale, and result in the computation of a set of scale values to represent the relative severity of the items. The remainder of this paper reports an experiment which demonstrates the validity of the ranking technique and hierarchical-structural design in measuring perceived crime severity.

5. THE EXPERIMENT

Kwan et al. (2000) (abbreviated as KIK hereafter) have used the Original Method to set up the seriousness scores for 15 crime typologies in Hong Kong. In the present experiment, a simple random sample of 200 students in

the Hong Kong Polytechnic University is to be drawn. Crime seriousness will be evaluated by both the Original Method and Simplified Method, and the results will then be compared with each other, and also with the KIK results. To reduce workload to an affordable level and avoid inconsistency of paired comparisons, the scope of the experiment is limited to eight crime typologies: indecent assault, rape, serious assault, burglary, theft, bribery, drug offense, and unlawful society. Brief descriptions of these crimes are given in KIK. In applying the Simplified Method, crimes are classified into three broad groups: burglary and theft are grouped into crime against 'Property'; bribery, drug offense and unlawful society into crime against 'Society'; and indecent assault, rape and serious assault into crime against 'Person'. The questionnaire consists of three sections: a section for making paired comparisons of the eight crimes, a section for ranking the severity of crimes within a group, and a section for ranking the three broad crime groups in terms of severity. Even in such a small experiment we can appreciate the workload reduction achieved. The Original Method requires 24 paired comparisons while the simplified version only requires 10 comparisons. The sequence of the sections in a questionnaire is rotated so as to avoid possible biases introduced by sequence of the questions. The fieldwork was conducted in November 2003. The results are presented below.

Table I presents a set of seriousness scores constructed by the Original Method for the eight crimes. The Simplified Method is used to construct first the proportions that a row crime group is more serious than a column crime group (Table II), then the proportions that a row crime is judged more serious than a column crime within a crime group (Tables III, IV and V), and finally the seriousness probabilities for the eight crimes (Table VI).

To assess the validity of the results of the Simplified Method, the seriousness probabilities from the Original Method in this survey and from the KIK study have been extracted for cross verification in Table VI. As KIK consider 15 crimes, their results have been recomputed using the paired comparison results of eight crimes that are considered here. Table VII gives the product moment correlations of the seriousness probabilities and the rank correlations of the seriousness ranks. Indeed a very high product moment correlation of 0.978 is reported between the seriousness probabilities obtained from the Original and the Simplified Methods, and a rank correlation of 0.929 is reported between their ranks. These indicate that the two methods yield very similar results. Moreover probabilities from the Simplified Method have a larger range (0.279), as compared with those from the Original Method (0.155). This indicates that the Simplified Method has better responsiveness than the Original Method in this experiment. The

TABLE I

Proportions column crimes judged more serious than row crimes, and seriousness scores using the Original Method

	Indecent assault	Rape	Serious assault	Burglary	Theft	Bribery	Drug offense	Unlawful society
Indecent assault		0.924	0.800	0.558	0.330	0.688	0.750	0.698
Rape	0.076		0.105	0.147	0.084	0.260	0.333	0.240
Serious assault	0.200	0.895		0.379	0.200	0.521	0.667	0.615
Burglary	0.442	0.853	0.621		0.124	0.742	0.670	0.649
Theft	0.670	0.916	0.800	0.876		0.773	0.794	0.742
Bribery	0.313	0.740	0.479	0.258	0.227		0.579	0.431
Drug offense	0.250	0.667	0.333	0.330	0.206	0.421		0.441
Unlawful society	0.302	0.760	0.385	0.351	0.258	0.569	0.559	
Column total	2.253	5.755	3.523	2.899	1.429	3.974	4.352	3.816
Seriousness probability	0.080	0.206	0.126	0.104	0.051	0.142	0.155	0.136

TABLE II

Proportions of column crime group judged more serious than row crime group using ranking data

	Property	Person	Society
Property		0.875	0.833
Person	0.125		0.438
Society	0.167	0.563	
Column total	0.292	1.438	1.271

TABLE III

Proportions of column crimes judged more serious than row crimes within 'Person' group using ranking data

	Indecent assault	Rape	Serious assault
Indecent assault		0.924	0.800
Rape	0.076		0.105
Serious assault	0.200	0.895	
Column total	0.276	1.819	0.905

TABLE IV

Proportions of column crimes judged more serious than row crimes within 'Property' group using ranking data

	Theft	Robbery
Theft		0.889
Robbery	0.111	
Column total	0.111	0.889

TABLE V

Proportions of column crimes judged more serious than row crimes within 'Society' group using ranking data

	Bribery	Drug offense	Unlawful society
Bribery		0.579	0.431
Drug offense	0.421		0.441
Unlawful society	0.569	0.559	
Column total	0.990	1.138	0.872

TABLE VI

Seriousness probabilities and ranks derived from the Original Method, the Simplified Method, and KIK

		Inde-cent assault	Rape	Serious assault	Bur-glary	Theft	Bribery	Drug offense	Un-lawful society
Seriousness probability	Original Method	0.080	0.206	0.126	0.104	0.051	0.142	0.155	0.136
	Simplified Method	0.044	0.291	0.145	0.085	0.012	0.140	0.161	0.123
	KIK	0.090	0.194	0.132	0.087	0.057	0.125	0.191	0.123
Seriousness rank	Original Method	7	1	5	6	8	3	2	4
	Simplified Method	7	1	3	6	8	4	2	5
	KIK	6	1	5	7	8	3	2	4

results give support to the validity of the Simplified Method. It should also be noted that although the KIK results came from a random sample of the Hong Kong population and the present experiment results from the

TABLE VII

Cross product-moment correlations of seriousness probabilities and cross rank correlations of seriousness ranks for the Original Method, the Simplified Method and the KIK

	Seriousness probability		Rank	
	Simplified Method	Original Method	Simplified Method	Original Method
KIK	0.904	0.930	0.976	0.905
Original Method	0.978	–	0.929	–

TABLE VIII

Serious scores by non-hierarchical design

	Indecent assault	Rape	Serious assault	Bur-glary	Theft	Bribery	Drug offense	Un-lawful society
Linear	8.046	20.551	12.585	10.351	5.103	14.194	15.542	13.628
Exponential	4.169	38.348	9.330	6.276	2.473	12.412	15.767	11.225
Information	7.777	21.329	12.471	10.130	4.856	14.193	15.661	13.583
Thurstone	8.621	19.802	12.461	10.879	6.108	13.849	14.918	13.362

Polytechnic University students, their product moment correlation and rank correlation were 0.930 and 0.905 for the serious probability and rank respectively, indicating that the students in the Polytechnic University and the Hong Kong population as a whole have similar attitude towards crime severity.

6. COMPUTATION OF SERIOUSNESS SCORES

We have computed the serious score for the two designs (hierarchical and non-hierarchical) using the Thurstone method. The results are given in the last rows of Tables VIII and IX, and are illustrated graphically in Figure 1. As the Simplified Method has a larger range in the serious probability it is also seen to yield a larger range in the serious score. This is expected due to the fact that scores are simply monotonically transformed from probabilities. Overall speaking, both methods lead to serious scores that are very close to each other. The Simplified Method which involves very much less workload than that of the Original Method can produce virtually the same scale values as the latter. This is also reflected by the high correlations between the serious probabilities (0.978) and between the serious scores

TABLE IX

Serious scores by hierarchical design

	Indecent assault	Rape	Serious assault	Bur-glary	Theft	Bribery	Drug offense	Un-lawful society
Linear	4.406	29.058	14.453	8.515	1.207	13.981	16.068	12.312
Exponential	1.873	62.767	7.838	3.364	1.188	7.328	9.865	5.777
Information	4.086	31.131	14.156	8.071	1.101	13.657	15.884	11.914
Thurstone	7.617	22.306	13.130	12.102	5.490	13.031	14.325	11.999

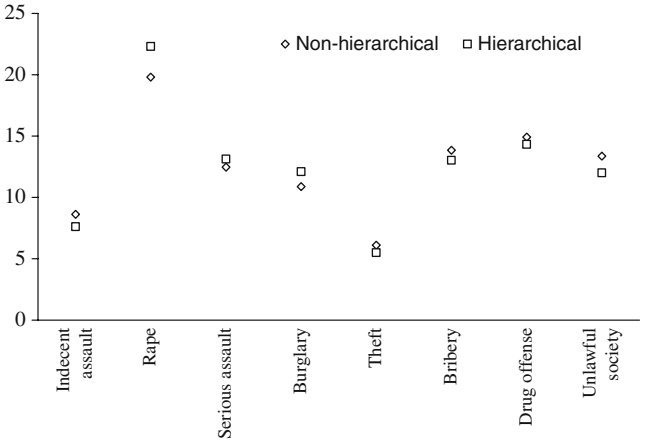

Fig. 1. Thurstone score.

(0.974) obtained by the two methods. Depending on the researcher's perception of the shape of the underlying continuum scale, the application of other models, such as linear, exponential and information, suggested by Ip et al. (2004) to obtain a new set of scale values is straightforward. The scores of these models are also reported in Tables VIII and IX for reference.

7. CONCLUSION

By organizing items into a hierarchical structure and by ranking the items and transforming them into paired comparisons, a lot of workload of data collection can be saved. Once the severity probability of items has been derived, any monotonic function will transform the severity probability to a continuum scale that preserves the severity ranking of items. The choice of model in measuring the relative importance of items depends very much on

researchers' perception of the shape of the underlying continuum of scale values, and on the results of validity and responsiveness checks.

ACKNOWLEDGEMENT

This research was supported by a research grant from The Hong Polytechnic University Research Committee.

REFERENCES

Bartholomew, D.J.: 1993, 'The statistical approach to social measurement', in D. Krebs and P. Schmidt (eds.), New Directions in Attitude Measurement (Walter de Gruyter, New York).

Bergner, M., R. Bobbitt, W. Carter and B. Gilson: 1992, Measuring Health: A Review of Quality of Life Measurement Scales (Open University Press, Buckingham).

Bowling, A.: 1997, Measuring Health: A Review of Quality of Life Measurement Scales (Open University Press, Buckingham).

Hunt, S.M., J. McEwen and S.P. Mckenna: 1986, Measuring Health Status (Croom Helm, London).

Ip, W.C., Y.K. Kwan, G.T.Y. Pong and M.D Trout: 1999, 'Development and demonstration of a hierarchical health status instrument for the elderly'. Journal of Social Service Research 25(4): pp. 43–56.

Ip, W.C., Y.K. Kwan and L.L. Chiu: 2004, Construction of health indices using paired comparisons. Social Indicators Research. To appear.

Kwan, Y.K., W.C. Ip and P Kwan: 2000, 'A crime index with Thurstone's scaling of crime severity'. Journal of Criminal Justice 28, pp. 237–244.

Kwan, Y.K., L.L. Chiu, W.C. Ip and P.Y.K Kwan: 2002, 'Perceived crime seriousness: Consensus and disparity'. Journal of Criminal Justice 30, pp. 623–632.

Lazarsfeld, P.F.: 1950, 'The logical and mathematical foundation of latent structure analysis', in S.A. Stouffer (ed.), Measurement and Prediction (Princeton University Press, Princton, NJ).

Medical Outcomes Trust: 1993, How to Score the SF-36 Health Survey (Collins, Boston).

Thurstone, L.L: 1927, 'A law of comparative judgement'. Psychometric Review 34, pp. 273–286.

Thurstone, L.L: 1928, 'Attitudes can be measured'. American Journal of Sociology 33, pp. 529–554.

Thurstone, L.L. and E.J. Chave: 1929, The Measurement of Attitude (University of Chicago Press, Chicago).

Ware, J.E. and C.D Sherbourne: 1992, 'The MOS 36-item short form health survey (SF-36): Conceptual framework and item selection'. Medical Care 30, pp. 473–483.

Department of Applied Mathematics
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
E-mail: mathipwc@polyu.edu.hk

W. C. Ip,
Y. K. Kwan,
L. L. Chiu