

FOUR KINDS OF REPRODUCIBILITY IN SCALE ANALYSIS¹

JOHN E. MILHOLLAND

University of Michigan

IN evaluating the adequacy of a unidimensional scale as a description of response patterns obtained from a set of items with two response categories, Guttman (2) uses what he calls a coefficient of reproducibility. It consists of the proportion of total responses which can be predicted from the scores, assuming that a unidimensional scale does in fact exist. For example, if there are 5 items, ordered 1, 2, 3, 4, 5, a person with a score of 4 should answer items 1, 2, 3, 4, positively; item 5, negatively. A pattern 1235 is considered as having one error, since changing the response to item 5 would change the pattern to 123, or changing the response to item 4 would change it to 12345, either of which is consistent with the scale. This method of counting errors gives a lower bound to the number of errors, since it is conceivable, in the example just given, that there may be errors in the responses to items 1, 2, and 3. What is done in every case to obtain an error score is to count the smallest number of responses which must be changed in order to make the pattern fit the scale. Reproducibility is then computed as the complement of the ratio of error score to total number of responses. In the example above, there is one error in five responses, so the reproducibility for a person giving the pattern is .80. For the group of individuals the computational process is similar, the error scores and total number of responses being summed over all individuals before the division is carried out.

The kind of coefficient just described makes the response of an individual to an item the unit of measurement of error. Three other different units might be used: an individual, a

¹ A preliminary version of this paper was read at the 1953 meeting of the Midwestern Psychological Association.

response pattern, and a response to an item in a response pattern, regardless of the frequency of occurrence of the response pattern. These four kinds of units lead to four corresponding reproducibility coefficients:

1. Items per individual. This is Guttman's coefficient described above.

2. Individuals. This is the proportion of individuals whose response patterns fit the scale perfectly. It is computed by dividing the number of such individuals by the total number of individuals.

3. Items per response pattern. In this instance, the frequencies of the various response patterns may be regarded as accidents of sampling. The mere existence of a response pattern entitles it to consideration on a par with all others. The coefficient is computed in the same fashion as in the case of the first coefficient, except that a response pattern is considered only once. The divisor of the fraction is the total number of items in the *different* response patterns.

4. Response patterns. This coefficient is the proportion of different response patterns which perfectly fit the scale. It is analogous to No. 2, except that, again, each response pattern is counted only once.

Three applications of these coefficients to data are given in Table I. The figures there were computed in a way slightly different from Guttman's, however. Individuals who responded to all items positively, or to all items negatively, were not included in the calculations. These people give no information about the structure of the scale (1)—they could be accommodated by any ordering of the items on a single dimension, or by any kind of multidimensional configuration, since they could be placed in the two opposite end cells in all dimensions.

The inclusion or exclusion of such cases does affect the reproducibility, of course. For example, on page 136 of *Measurement and Prediction* (4) there is a scalogram pattern with reproducibility .98. Of the hundred cases in the sample, however, 45 responded either positively to all items or negatively to all items. When these 45 cases are taken out, the remaining 55 exhibit a reproducibility of .88, which is the figure given in the table under coefficient No. 1. For the purpose of comparing the

TABLE 1
Reproducibility Coefficients Computed by Four Different Methods

Data and Source	N	Reproducibility*			
		1.	2	3	4
Attitude toward army career, 6 items (4) (For $N = 100$, reproducibility No. 1 is .98)	55				
One dimension		88	80	85	42
Two dimensions		99	96	97	83
Fear symptoms, 9 items (4) (For $N = 100$, reproducibility No. 1 is .92)	87				
One dimension		91	45	87	17
Three dimensions		96	72	94	54
Strong Vocational Interest Blank, 13 items (3)	729				
Five dimensions		95	57	94	41
Six dimensions		97	67	95	54
Eight dimensions		99	82	98	73

* Computation formulas Key: D = total number of different response patterns
 E = number of individual error responses

1. $\frac{nN - E}{nN}$
2. $\frac{I}{N}$ e = number of error responses in the *different* response patterns
3. $\frac{nD - e}{nD}$ I = number of individuals whose response patterns fit the scale perfectly
4. $\frac{R}{D}$ N = number of individuals
 n = number of items
 R = number of different response patterns which fit the scale perfectly

adequacy of different scales or different scale structures, it would seem desirable to make the comparisons on the basis of individuals for whom the various alternative solutions would make a difference.

For the example just cited, it can be seen by reference to the table that the first three coefficients are relatively high, but that less than half of the *response patterns* can be fitted to a unidimensional solution. If, however, two dimensions are admitted², each being a Guttman-type scale, 83 per cent of the response patterns can be fitted into cells of the two-dimensional space.

Each of the reproducibility coefficients would seem to be a useful and different index of the extent to which a proposed configuration accounts for the observed data. If the concern is

² The method for obtaining the multidimensional configuration is described in (3).

with the description of individuals, coefficient No. 2, which tells the proportion of individuals whose responses can be completely specified by the configuration, is probably the most meaningful. A comparison of this coefficient with coefficient No. 1, items per individual, gives a measure of how extensive the failure to fit the residual group of individuals is. If the two coefficients are nearly equal, it means that very few of the cases not fitted perfectly are even partially fitted, since precise equality of the coefficients would mean that all the responses of the people in the residual group were error responses. Coefficient No. 2 is, of course, a lower bound for No. 1, and if No. 1 is considerably the higher of the two it signifies that the configuration closely approaches fitting the response patterns which are not fully accommodated.

If the primary problem of a study is the analysis of the multidimensional structure of a set of items, proportionately more emphasis may be given to obtaining a high value for the fourth coefficient, response patterns. The very fact that a response pattern does occur testifies to its possibility of occurrence, and a measure of the extent to which a solution is providing for all cells of the multidimensional space is given by this coefficient. The third coefficient, items per response pattern, assesses the deviations of the patterns not fitted in a manner analogous to the way No. 1 performs for individuals, as described in the preceding paragraph. Likewise, No. 4 is a lower bound for No. 3.

In most situations it is to be expected that an attempt will be made to have the configurations fit the more frequently oc-

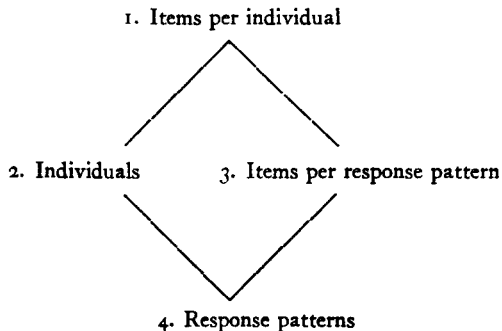


FIG. 1. Relations Among the Magnitudes of the Four Reproducibility Coefficients.

curing response patterns in preference to the rarer ones, in which case the relationships among the magnitudes of the various coefficients may be represented by the partial order shown in Figure 1.

The circumstances that coefficient No. 1 is larger than No. 2, and that No. 3 is larger than No. 4 have been pointed out previously. The relations shown in the diagram between No. 1 and No. 3, and between No. 2 and No. 4 are based on conjecture, but No. 1 will be larger than No. 3 if the more frequently appearing response patterns are the more closely fitted; and No. 2 will be larger than No. 4 if the patterns having the larger frequencies are the ones which are perfectly fitted.

REFERENCES

1. Edwards, A. L. "On Guttman's Scale Analysis." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VIII (1948), 313-318.
2. Guttman, L. "The Basis for Scalogram Analysis." In S. A. Stouffer et. al., *Measurement and Prediction*. Princeton: Princeton University Press, 1950.
3. Milholland, J. E. "The Dimensionality of Response Patterns for the Method of Single Stimuli." Unpublished doctor's dissertation. Ann Arbor: University of Michigan, 1953.
4. Suchman, E. A. "The Utility of Scalogram Analysis." In S. A. Stouffer et. al. *Measurement and Prediction*. Princeton: Princeton University Press, 1950.

VALIDITY STUDIES SECTION

Edited by
WILLIAM B. MICHAEL