Item Homogeneity Defined by Multivariate Symmetry

Gustav A. Lienert
University of Erlangen-Nürnberg

Ulrich Raatz
University of Duisburg

The concept of item homogeneity, implying that items have equal difficulties and equal intercorrelations, is defined by multivariate axial symmetry. A nonparametric test for J-item homogeneity is proposed and illustrated by a numerical example.

Item homogeneity is intuitively understood as the degree to which a sample of J items measures the same concept (intelligence or neuroticism) along a one-dimensional scale. Operationally, item homogeneity can be defined by inter-item correlations and evaluated by factor analysis: If factor analysis of J items results in one common factor, the items are said to be homogeneous (Lord & Novick, 1968, p. 381). A more liberal definition requires that those items that constitute a one-dimensional space are said to be homogeneous, even though they may be composed of more than one common factor.

Both definitions of item homogeneity are parametric definitions, implying that correlations between pairs of items, say the ($\frac{4}{2}$) phi coefficients, exhaust all information on the multivariate contingency among the J items.

Testing item homogeneity as defined in this way is important for constructing one-dimensional test scales by counting the number, x, of binary items in the keyed direction.

This paper will propose a nonparametric definition of item homogeneity and a test for it based on a multivariate generalization of McNemar's (1947) test of bivariate symmetry in a fourfold table.

A Generalized Definition of Item Homogeneity by Multivariate Axial Symmetry

The strong definition of item homogeneity as a set of J items with equal loadings on a common factor implies that the items are of equal difficulty and have equal intercorrelations (phi coefficients). Equal difficulty indices and equal phi intercorrelations, however, are given if the J items constitute a 2^J-dimensional contingency cube with J-variate axial symmetry, as shown numerically in Figure 1 by perfect trivariate axial symmetry.

For example, suppose a sample of $n = 60$ testees was given a three-item test. Each individual testee produces the answer pattern (configuration)

$$C = (i_1, i_2, i_3) \qquad [1]$$

in Yes-No answers to the three items where $i_1 = 0$ or 1; $i_2 = 0$ or 1; and $i_3 = 0$ or 1, with $i_j = 0$ for a No answer and $i_j = 1$ for a Yes answer to item j.

The observed frequencies of the 2^3 answer patterns, C, are presented in the three-dimensional contingency cube of Figure 1 with one dimension for each item. Thus, $n_{111} = 17$ testees had the answer pattern (1, 1, 1), i.e., (Yes, Yes, Yes) and so forth. The contingency cube with observed frequencies found in Figure 1 shows perfect trivariate axial symmetry of the answer patterns as defined by Wall (1976): If the frequencies n_{000} and n_{111} of the symmetry axis are ignored (such as n_{00} and n_{11} in a

Figure 1
$J = 3$ Items which are Trivariate Axial Symmetric
with Equal Intercorrelations in Their Two-Dimensional Marginals
and Equal Difficulties in Their One-Dimensional Marginals,
with $n = 66$ Testees
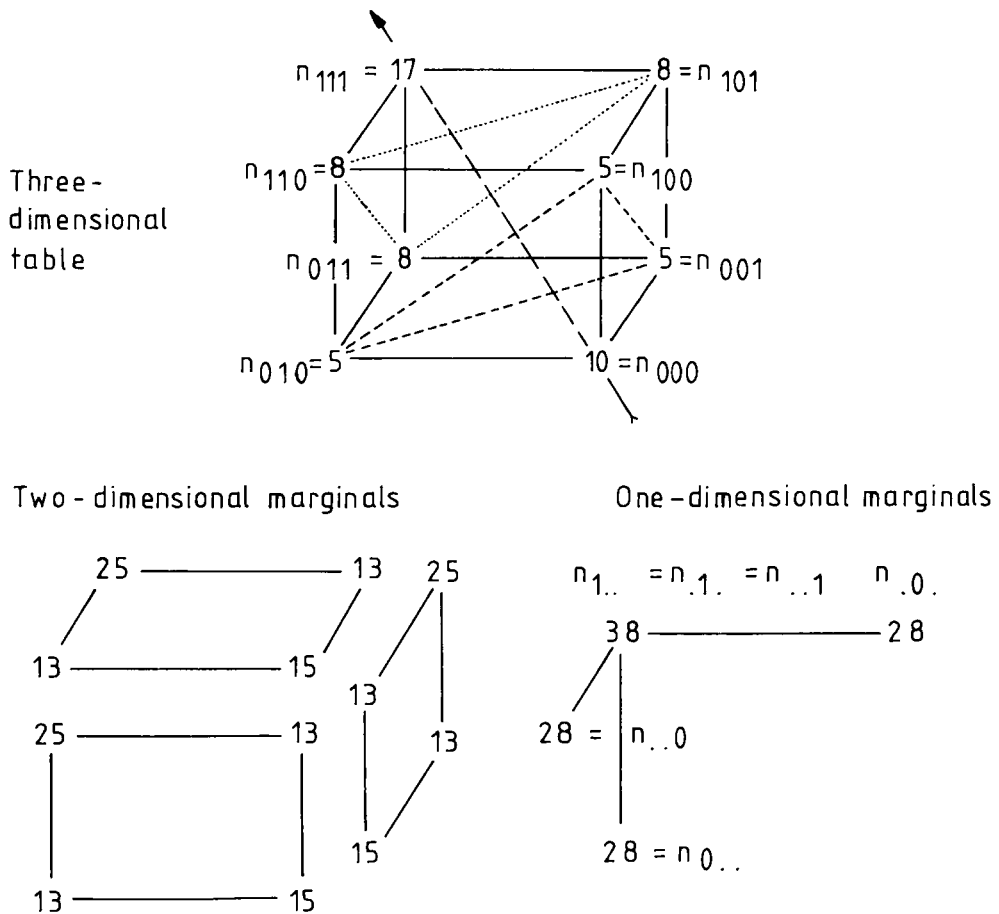
McNemar table), the frequencies of each of the two triangles in Figure 1 are identical. Any one triangle encloses all answer patterns with the same number, x, of Yes answers: There are $n_{100} = n_{010} = n_{001} = 5$ testees who have one of the three answer patterns with $x = 1$ Yes answers and $n_{110} = n_{101} = n_{011} = 8$ testees who have one of three answer patterns with just $x = 2$ Yes answers.

Thus, it is evident from the one-dimensional marginals of the contingency cube in Figure 1 that all items have difficulties of $n_1 /n = n_{\cdot 1} /n = n_{\cdot\cdot 1}/n = 38/66 = .576$, where $n_1 = n_{111} + n_{110} + n_{101} + n_{100} = 17 + 8 + 8 + 5 = 38$ is the number of testees who answered Item 1 with Yes.

Looking at the two-dimensional marginals of the contingency cube in Figure 1, it is evident that they are composed of identically distributed and McNemar-symmetric fourfold frequencies, corresponding to phi intercorrelations of $\phi = (25 \times 15 - 13 \times 13) / \sqrt{38 \times 28 \times 38 \times 28} = 194$, where $n_{11} = n_{111} + n_{110} = 17 + 8 = 25$, $n_{10\cdot} = n_{101} + n_{100} = 8 + 5 = 13$, $n_{01} = n_{010} + n_{011} = 5 + 8 = 13$ and $n_{00} = n_{000} + n_{001} = 10 + 5 = 15$ are the fourfold marginal frequencies of the three item intercorrelations.

Thus, trivariate axial symmetry implies that the three Yes-No, or binary, items have equal difficulty indices of $p = .576$ and equal phi intercorrelations of $\phi = .194$, as shown by trivariate axial symmetry constructed artificially. In real sampling, of course, trivariate symmetry will deviate more or less from ideal symmetry, and the question is how to test whether trivariate, or multivariate, axial symmetry exists in the population from which the sample of n testees has been gathered. A large-sample answer to this question will be given below.

A Chi-Square Test of Multivariate Axial Symmetry in a 2^J Contingency Table with J Binary Items

Assume that J binary items have been given to a random sample of n testees, where each testee responds to each item with Yes (1) or No (0). The null hypothesis is that the J items form a J-dimensional contingency cube, which is axially symmetric for J. The omnibus alternative hypothesis is that the J items are not axially symmetric, i.e., that they are not homogeneous according to the new definition. The rationale for testing axial symmetry, i.e., homogeneity of J items may be derived from specifying Wall's (1976) concept of axial symmetry of multinary variables to binary variables as follows:

Assume that an individual testee produces the answer pattern

$$C = (i_1, i_2 \ldots\ldots, i_j, \ldots, i_J) \qquad [2]$$

in responding to J items, where $i_1 = 0$ or 1; $i_2 = 0$ or 1; and so forth to $i_J = 0$ or 1, with $i_j = 0$ for a No answer and $i_j = 1$ for a Yes answer to item j. Then, $n_x (i_1, i_2, \ldots, i_J)$ denotes the frequency with which the answer pattern $(i_1, i_2, \ldots, i_J)$ has been observed in the subsample of n_x testees who gave exactly x Yes answers to the J items.

H_0 now states that the probabilities $p_x (i_1, i_2, \ldots, i_J)$ of all patterns involving exactly x Yes answers are equal. S_x denotes the set of all answer patterns $(i_1, i_2, \ldots, i_J)$ containing exactly x Yes answers, and H_0 is defined by

$$H_O : p_x (i_1, i_2 \ldots\ldots, i_J) = \text{const} = p_x \qquad [3]$$

where $x = 0 (1) J$ are the test scores in the conventional concept of cumulative scoring and $x = 1 (1) J - 1$ are the scores which are of interest for testing H_0.

Since, in general, the probability parameters p_x are unknown for a sample of n testees, they must be estimated by imposing the restriction of a fixed score distribution onto the sample of n testees by setting

for all $(i_1, i_2 \ldots, i_J) \; \varepsilon \; S_x$

$$\sum_{(i_1, i_2, \ldots, i_J) \; S_x} n_x (i_1, i_2, \ldots, i_J) = n_x \qquad [4]$$

fixed for all x

Thus, the probability estimate n_x of a particular answer pattern $(i_1, i_2, \ldots, i_J)$ from S_x depends only on the number Z_x of elements in the set S_x of possible answer patterns. Since Z_x is the number of permutations of $(i_1, i_2, \ldots, i_J)$ without repetition, it is given by the binomial coefficient

$$Z_x = \binom{J}{x} = \frac{J!}{x! \; (J - x)!} \qquad [5]$$

Under H_0 with the above restriction, the expected frequency of an observed frequency $n_x (i_1, i_2, \ldots, i_J)$ in the subsample of testees with exactly x Yes answers is given by

$$\hat{e} (i_1, i_2, \ldots, i_J) = n_x (i_1, i_2, \ldots, i_J) / Z_x = : \hat{e}_x \qquad [6]$$

$$\hat{p} (i_1, i_2, \ldots, i_J) = \hat{e}_x / n \qquad [7]$$

can be shown to be maximum likelihood estimate of the probability parameter p_x.

Thus, for sufficiently large e's, axial symmetry for all $\binom{J}{x}$ answer patterns with exactly x Yes answers may be evaluated by chi-square, i.e., by means of

$$X_x^2 = \sum_{S_x} \frac{\left| n_x (i_1, i_2, \ldots, i_J) - \hat{e}_x \right|^2}{\hat{e}_x} \qquad [8]$$

where X_x^2 is distributed approximately as χ^2 with $Z_x - 1$ *df* under H_0 of x Yes answers or x-score symmetry.

A test of axial symmetry for all J items, including all possible x-score symmetries, is given by

$$X^2 = \sum_{x-1}^{J-1} X_x^2 \qquad [9]$$

since the $J - 1$ components of X^2 are independent of each other, i.e., made up from independent subsamples of n_x testees each. Thus, the global test statistic X^2 of multivariate axial symmetry of all J test items is distributed as χ^2 with $\Sigma(Z_x - 1)$ *df*'s, where summing is from $x = 1$ to $x = J - 1$, since for $x = 0$ and for $x = J$, the observed frequencies are identical to the expected frequencies and thus give 0 chi-square components. Note that the same is true for McNemar's test where the cell with $x = 0$ and the observed frequency n_{00}, and the cell with $x = J = 2$ and the observed frequency n_{11} do not contribute to $\chi^2 = (n_{10} - n_{01})^2 / (n_{10} + n_{01})$. In every case the frequencies of the two cells that constitute the axis of bivariate or multivariate symmetry are ignored in symmetry testing.

Using Equation 9, H_0 of multivariate axial symmetry of J binary items must not be rejected if χ^2 does not exceed the $(1 - \alpha)$ quantile χ_{df}^2 of the χ_{df}^2 distribution. In this case the J items may assumed to be homogeneous in the sense defined.

A Numerical Example

$J = 4$ items drawn from an intelligence test (DST; Lienert, 1964) are given to a sample of $n = 94$ students. The question of interest is whether the four items may be said to be homogeneous.

The null hypothesis of item homogeneity is assumed to hold if the four items have a symmetrical distribution along the main diagonal of the 2^4 contingency table. Since all expected frequencies under H_0 are larger than two, the asymptotic test described may be used for testing against item homogeneity (multivariate asymmetry).

The answer patterns and their observed frequencies n are listed in Table 1. In Table 1, x is the number of Yes answers in the pattern (i_1, i_2, i_3, i_4) and Z_x is the number of different patterns with x Yes answers; $n_x = \Sigma f$ is the observed frequency of patterns with x Yes answers, and $\hat{e}_x = n_x/Z_x$ is the frequency of every pattern expected under H_0 of four-item homogeneity. X_x^2 is the component resulting from the homogeneity of the patterns with exactly x Yes answers. Their sum is distributed as chi-square, which is not significant at the 5% level for $df = 5 (Z_x - 1) = 2^4 - 5 = 11$ df, where 5 is the range of a 4-point scale. Consequently, the $J = 4$ items may be assumed to be homogeneous.

From Table 1 the difficulty indices and the phi-coefficients can be calculated. The difficulty index of Item 1 is, e.g., $p_1 = n_{1...}/n = 8+7+5+6+5+4+3+6/94 = .47$. The four frequencies for the fourfold

Table 1
Frequency of Answer Patterns of the $n = 94$ Testees of the Example, and the Calculation of Chi-Square.

y	$i_1 i_2 i_3 i_4$	f	Z_x	n_x	$\hat{e}_x$	$(f - \hat{e}_x)^2/\hat{e}_x$	χ_x^2
0	0000	13	1	13	13.00	.000	
1	1000	8	4	25	6.25	.490	.760
	0100	5				.250	
	0010	6				.010	
	0001	6				.010	
2	1100	7	6	36	6.00	.167	.667
	1010	5				.167	
	1001	6				.000	
	0110	6				.000	
	0101	7				.167	
	0011	5				.167	
3	1110	5	4	14	3.50	.643	1.428
	1101	4				.071	
	1011	3				.071	
	0111	2				.643	
4	1111	6	1	6	6.00	.000	

$$n = 94 \qquad df = 16 - 5 = 11 \qquad \chi^2 = 2.855$$

table for the correlation, e.g., of the Items 1 and 2 are n_{00}, n_{01}, n_{10} and n_{11}, where, e.g., n_{00} = 13+6+6+5 = 30, and so forth.

Note that the difficulty indices of the four items ($p_1 = .47$, $p_2 = .45$, $p_3 = .40$, $p_4 = .42$) calculated from Table 1 are quite similar in size. The same is true for the six phi intercorrelations ($\phi_{12} = .10$, $\phi_{13} = .05$, $\phi_{14} = .03$, $\phi_{23} = .09$, $\phi_{24} = .07$, $\phi_{34} = .02$) calculated from Table 1. These phi coefficients are close to zero, which indicates that the four problems, though homogeneous in the symmetry definition, measure different aspects of problem-solving behavior in the 94 testees.

Grouping Items for Homogeneity

Assume that the test constructor has preselected by content analysis J items that he/she found to be nonhomogeneous by symmetry testing in a sample of n individuals. How should he/she proceed in such a situation to achieve homogeneity in a smaller set of J_1 items?

One procedure is to calculate the difficulty indices of every item and the phi coefficients for every pair of items. Excluding the item most deviant in difficulty or most deviant in its intercorrelations with the remaining $J - 1$ items will perhaps lead to a nonsignificant chi-square.

If excluding one single item is not successful, a heuristic grouping procedure may be tried out to obtain subscales of homogeneous items as follows: If only two subgroups are admitted, subdivide the J items into J_1 and $J - J_1$ items in every possible way and accept the grouping that results in a nonsignificant symmetry chi-square for each subgroup. Each subgroup of items then constitutes a separate subscale.

If there is no subdivision that gives two homogeneous subgroups of J_1 and $J_2 = J - J_1$ items, the next step is to try out a three-group subdivision with J_1, J_2, and J_3 homogeneous items. If at least one of the possible $J!/J_1!J_2!J_3!$ groupings results in three nonsignificant symmetry chi-squares, the test constructor has arrived at a test battery with three homogeneous subscales. Note that neither item exclusion nor grouping protects the alpha risk of symmetry testing.

Conclusion

Multivariate axial symmetry seems to be most useful in constructing tests with few but complex items, such as problem-solving tests. Tests of this type are most informative if the J items have low intercorrelations covering broad constructs, such as problem-solving power. Care should be taken that only few testees give all Yes or all No answers to the J items to be scored. This goal is best achieved by selecting items of difficulties close to .5.

In constructing J parallel forms of a conventional test, the symmetry test may be used to establish the equivalence of J-tuplets of items with the same content. Though the items of a conventional test may vary in their difficulty indices, the parallel items must fulfill the symmetry requirement if the J parallel tests are strictly parallel.

If descriptive evaluation of homogeneity is desired in large samples of testees, use Pearson's contingency coefficient or its square

$$c^2 = \chi^2 / (n + \chi^2) \hspace{2cm} [10]$$

as a measure of the nonhomogeneity of J items.

In practice, the main objection against using the multivariate axial symmetry test for test construction is the large sample size required, even for small sets of binary items. Note that test scales constructed by sampling J homogeneous items have, in general, ordinal rather than interval char-

acteristics. Only if the J items are independent of each other, and if the test scores fit a binomial distribution, may the resulting scale claim to be of interval level.

The symmetry test described in this paper is a generalization of the McNemar (1947) test limited to binary items (Yes-No). But often multinary items (Yes, Don't Know, No) are given. In this case, the two-dimensional Bowker symmetry test (1948) may be generalized in the same manner as is done in the approach used in this paper. A symmetry concept such as the one proposed here for defining the item homogeneity of more than two items has not, as yet, been cited in discrete multivariate test books (e.g., Bishop & Fienberg, 1975 and Fienberg, 1979). The question is how axial symmetry may be related to the traditional parametric concept or to other concepts.

If axial symmetry is given, all item intercorrelations are equal. This implies a general factor with equal loadings on the items according to the one-factor homogeneity definition given by Lord and Novick (1968, p. 536). Nevertheless, axial symmetry further deserves a comparison with factor analysis beyond the level of this short discussion.

The approach presented here is made in the traditional frequency model but may be related to the more modern log-linear model of contingency analysis (see Fienberg, 1977, chap. 3). Complete axial symmetry implies equal main effects and equal first-order interaction effects in the corresponding log-linear model. On the other hand, the null hypotheses of no main effects and no interaction effects in the log-linear model correspond to a special axial symmetry with items of middle difficulty and zero intercorrelations. Further discussion is needed, however, for this problem.

References

Bishop, Y. M., & Fienberg, S. E. *Discrete multivariate analysis: Theory and practice.* Cambridge MA: MIT Press, 1975.

Bowker, A. H. A test for symmetry in contingency tables. *Journal of the American Statistical Association.* 1948, *43*, 572-74.

Fienberg, S. The analysis of cross-classified categorical data. Cambridge MA: MIT Press, 1979.

Lienert, G. A. *Denksporttest (DST).* Göttingen: Hogrefe, 1964.

Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores.* Reading MA: Addison-Wesley, 1968.

McNemar, Q. Note on sampling error of the differences between correlated proportions or percentages. *Psychometrika,* 1947, *12*, 153.

Wall, K. D. Ein Test auf Symmetrie in einer J-dimensionalen Kontingenztafel. *EDV in Biologie und Medizin,* 1976, *7,* 57-64.

Authors' Address

Send requests for reprints or further information to either Prof. Dr. G. A. Lienert, Universität Erlangen—Nürnberg, Lehrstuhl für Psychologie, Regensburger Str. 160, D-8500 Nürnberg 30, Federal Republic of Germany or Prof. Dr. U. Raatz, Universität Duisburg, FB 2—Psychologie, Lotharstr. 65, D-4100 Duisburg, Federal Republic of Germany.