DOCUMENT RESUME

ED 117 124 TM 004 960

AUTHOR	Cliff, Norman
TITLE	A Basic Test Theory Generalizable to Tailored Testing. Technical Report No. 1.
INSTITUTION	University of Southern California, Los Angeles. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
REPORT NO	TR-1
PUB DATE	22 Oct 75
NOTE	50p.
EDRS PRICE	MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS	*Computer Oriented Programs; *Individual Differences; Item Analysis; *Item Sampling; Matrices; Response Style (Tests); *Testing; *Test Reliability
IDENTIFIERS	Ordering Theory; *Tailored Testing; Test Theory

ABSTRACT

Measures of consistency and completeness of order relations derived from test-type data are proposed. The measures are generalized to apply to incomplete data such as tailored testing. The measures are based on consideration of the items-plus-persons by items-plus-persons matrix as an adjacency matrix in which a 1 means that the row element dominated the column element regardless of whether it is item or person which dominates. From this the number of item-item and person-person dominance relations can be derived. Indices of consistency are developed which compare the number of dominance relations in a matrix to the number that would occur if it were perfectly consistent and to the number in a random matrix. The indices of consistency are shown to relate to test variance, the Kuder-Richardson formulas, and Loevinger's homogeneity indices. Indices of completeness are also proposed. (Author)

A BASIC TEST THEORY GENERALIZABLE TO TAILORED TESTING

Norman Cliff

Technical Report No. 1

Department of Psychology
University of Southern California
Los Angeles, California 90007

October, 1975

Prepared under contract No. N00014-75-C-0684
NR No. 150-373, with the Personnel and
Training Research Programs, Psychological Sciences Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A BASIC TEST THEORY GENERALIZABLE TO TAILORED TESTING		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Norman Cliff		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0684
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS G1153N, RR042-04, RR042-04-01, NR150-373
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, Virginia 22217		12. REPORT DATE October 22, 1975
		13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

(U) Tailored Testing; (U) Test Theory; (U) Consistency; (U) Guttman Scaling; (U) Implied orders

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Measures of consistency and completeness of order relations derived from test-type data are proposed. The measures are generalized to apply to incomplete data such as tailored testing. The measures are based on consideration of the item-plus-persons by items-plus-persons matrix as an adjacency matrix in which a 1 means that the row element dominated the column element regardless of whether it is item or person which dominates.

From this the number of item-item and person-person dominance relations

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 (cont.)

can be derived. Indices of consistency are developed which compare the number of dominance relations in a matrix to the number that would occur if it were perfectly consistent and to the number in a random matrix. The indices of consistency are shown to relate to test variance, the Kuder-Richardson formulas, and Loevinger's homogeneity indices. Indices of completeness are also proposed.

A Basic Test Theory Generalizable to Tailored Testing

The available methods or strategies for tailored testing seem not to include provisions for a self-contained evaluation of the test. They seem to make little provision for estimating anything analagous to a Kuder-Richardson (Gulliksen, 1950) or alpha (Cronbach, 1950) reliability of the instrument itself in the tailored context. Rather, they assume the availability of precise item statistics derived from other sources. The basic obstacle to deriving such consistency information comes from the difficulty of handling the incomplete nature of the item response data in the tailored or computer interactive testing case. An attempt will be made here to provide ways of filling this need by suggesting a way to conceptualize the system.

There is a point of view which looks on testing as an ordering process, ordering persons with respect to each other or ordering persons relative to items or tasks. The fact that what Coombs (1964) calls "a dominance relation between members of different sets" can lead to a formally consistent ordering has been known for a considerable time. Guttman (1941, 1950) was the major pioneer, but the work of Loevinger (1947, 1948) repeated and expanded on many of those principles.

Recently, there has been a revived interest in ordinal testing concepts, e.g., Ducamp and Falmagne (1969), Airasian and Bart (1973), Bart and Krus (1973), Krus (1974), Cliff (1975). These papers include reformulations and elaborations on the use of binary (right-wrong) information to develop orders, and they lead to the formulations presented here. It can be shown that a Guttman scale leads to a joint partial-order of persons and items. Exact Guttman scales--like other partial orders--rarely exist in nature, and then only in artificial environments, but approximations to them are not uncommon. The degree to which a given matrix of relations resembles a partial order can be measured, and these measures will be proposed as indices of the consistency of measurement of a test. In the case of complete data, some of these measures will be found to be related to the familiar Kuder-Richardson (KR) formulas (Gulliksen, 1950) and to some proposed by Loevinger (1947). New interpretations of those formulas result. Some of the measures will be new, and it will be contended that they are more relevant than the traditional ones, as measures of order.

The formulations lead quite directly to generalizations which are applicable to the incomplete data--tailored or computer-interactive--case. That is, measures of the internal consistency of responses will be developed which are entirely self-contained in the tailored testing process rather than relying on outside information.

Notation and Basic Principles

The basis of recent developments in tests as ordering instruments is the consideration of the score matrix as items-plus-persons by items-plus-persons rather than simply items by persons. Then a 1 indicates a dominance relation of row element over column element, regardless of whether item or person is dominant. The following conventions are used in later developments.

The matrix S contains item-person responses with $s_{ij} = 1$ if person i answers item j correctly; $i = 1, 2, \ldots n$; $j = 1, 2, \ldots x$. $\bar{S}$ is a matrix of item-person responses with $\bar{s}_{ij} = 1$ if person i answers item j *incorrectly*.

A is a supermatrix:

$$(1) \qquad A = \begin{bmatrix} 0 & \tilde{S}' \\ S & 0 \end{bmatrix}$$

where 0 indicates null sections and $\tilde{S}'$ is the transpose of $\tilde{S}$. Thus the elements of A indicate dominance relations of row element over column element, regardless of whether the elements are items or persons. Item-item and person-person relations are always 0-0 because there is no direct observation of these relations.

For complete data, S and $\tilde{S}$ are complements,

$$(2) \qquad s_{ij} = 1 - \tilde{s}_{ij}.$$

In matrices

$$(3) \qquad S = \underset{\sim}{1}\,\underset{\sim}{1}' - \tilde{S},$$

where $\underset{\sim}{1}$, as elsewhere below, denotes a column of unities. The complementary relation does not hold in tailored testing where some elements are zero in both matrices because the person was not presented the item.

The ordinary matrix product of A with itself is denoted

$$(4) \qquad A^2 = \begin{bmatrix} N & 0 \\ 0 & X \end{bmatrix}$$

Note that all even powers of A have the general form of (4) whereas the odd powers resemble (1). The scalar notation a_{ijp} denotes the ij element of A^p. In (4), where

$$(5) \qquad N = \tilde{S}'\, S,$$

the element n_{kj} is equal to the number of persons who get k wrong and j right, i.e., the number of times item k dominates j. Similarly

(6) $X = S\tilde{S}'$

is a person-dominance matrix, with element x_{ih}, and contains the number of times person i dominates person h, i.e., the number that i gets right and h wrong.

Total scores are also used:

(7) $X_i = S1$ and

(8) $\tilde{X}_i = \tilde{S}1$

Similarly,

(9) $N_j = 1'S$ and

(10) $\tilde{N}_j = 1'\tilde{S}$.

The vector N_j should be noted as a difficulty score, the number of item "wins," not the number passing as is more typically used.

For complete data,

(11) $x_i = x - \tilde{x}_i$, and

(12) $n_j = n - \tilde{n}_j$,

but again the relations do not hold for the missing data case.

Use is made of the "Boolean powers" of A as well as the numerical ones, particularly in considering the logical properties of the relations. Here, the matrix multiplication uses Boolean arithmetic in which $1 + 1 = 1$, $1 + 0 = 0 + 1 = 1$, $0 + 0 = 0$; $1 \cdot 1 = 1$, $0 \cdot 1 = 1 \cdot 0 = 0$, $0 \cdot 0 = 0$. Such powers are denoted by having the exponent in parentheses

(13) $$A^{(2)} = \begin{bmatrix} D & 0 \\ 0 & Y \end{bmatrix}$$

where D and Y are the logical counterparts of N and X, respectively.

Still a third form of powers of A is used, A^{p*} in which binary entries record the relation between the elements of A^p and some threshold values for them. Let Z^p be an n + x by n + x matrix of threshold or criterion values for A^p. Then

$$(14) \qquad A^{p*} = (A^p > Z^p).$$

That is, $a^*_{ijp} = 1$ if $a_{ijp} > z_{ijp}$ and zero otherwise. On occasion we may have cause to refer to the submatrices of A^{p*} as S*, $\tilde{S}$*, N*, or X* and to their elements s^*_{ij}, etc.

Thus $A^{(p)} = A^{p*}$ if $Z^p = 0$. The matrix Z^p could also be a constant such as 2.0 if, for example, it were felt that at least two connecting relations should be necessary to imply that i dominates h. Krus (1974) in effect defines z_{ihp} as the number of relations necessary to show that i "significantly" dominates h.

The foregoing are the main notational conventions that are used below. A few additional ones are introduced as needed.

Indices of Orderliness

Person i "dominates" person h when he gets an item right which h gets wrong. Item j dominates item k when there is a person who gets j wrong and k right. The dominance relations are consistent insofar as they occur in only one direction, i.e., are asymmetric rather than there being instances of both j dominating k and the reverse. This is what is meant by reproducibility in a Guttman scale (Guttman, 1950). If the objective of the measurement process is to order persons or items or both, this ordering should be complete as well as consistent, at least in most instances, for a highly incomplete ordering is almost as useless as a highly inconsistent one. Consequently, indices of completeness as well as consistency are called for.

A useful consequence of the use of the complete matrix A rather than simply the rights matrix S is that every index derived for items implies a parallel one for persons. This is a complete duality as substitution of items for persons and vice versa in any formula or theorem will result in a legitimate parallel formula or theorem. Indices derived for items will often turn out to be related to traditional statistics and formulas although it will often be argued that the parallel ones for persons are actually more relevant to the purposes of testing.

The number of dominance relations

For the most part we will concentrate on equations relating to the numerical mode of operation implied by Equations 4, 5, and 6 although the more qualitative formulations of Equations 13 and 14 will be returned to later so that some analagous measures can be proposed.

The primary dominance information is simply the number of elements in the person matrix N and the item matrix X. We denote these quantities v and g, respectively:

(15) $v = \sum_{j}\sum_{k} n_{jk}$ and

(16) $g = \Sigma\Sigma x_{ih}$

In matrices, these are

(17) $v = 1'(S'S)1$, and

(18) $g = 1'(SS')1$,

respectively, so that relocation of the parentheses makes it clear that

$$(19) \qquad v = \sum_i \alpha_i x_i \quad \text{and}$$

$$(20) \qquad g = \sum_j \alpha_j n_j \ .$$

Minimum number of dominance relations

Now suppose the sum takes place only above the diagonals of the matrices:

$$(21) \qquad v_m = \sum_j \sum_{k>j} n_{jk} \ ,$$

$$(22) \qquad g_m = \sum_i \sum_{h>i} x_{ih} \ .$$

If the items or persons are in the appropriate order, and the data are perfectly consistent, then the matrices will be upper triangular, and v will equal v_m and g will equal g_m, otherwise they will be larger. This suggests the ratios g_m/g and v_m/v as consistency indices, but they have the drawback that they will tend to be fairly close to unity for most data, as the coefficient reproducibility is. They do, however, have the advantages over the latter of having terms with a clear interpretation and being straightforward to compute.

The indices v_m and g_m can be computed from the marginals without reference to the n_{jk} at all. If the data are consistent, $n_{kj} = 0$ for $k > j$. Subtracting n_{kj} from each term of (21) gives

$$(23) \qquad v_m = \sum_j \sum_{k>j} (n_{jk} - n_{kj}) \ .$$

But $n_{jk} - n_{kj} = n_j - n_k$, so

$$(24) \qquad v_m = \sum_j \sum_{k>j} (n_j - n_k) \ .$$

This reduces to

$$(25) \qquad v_m = \sum_j (n - 2_j + 1) n_j ,$$

or

$$(26) \qquad v_m = (x + 1)\Sigma n_j - 2\Sigma j n_j$$

Similarly,

$$(27) \qquad g_m = \Sigma (x - 2i + 1)\, x_i$$

Thus, it is not necessary to actually have the N or X matrices to compute v, v_m, g, or g_m.

An obvious correction term to apply to g_m/g is to assume that there is no order, everything is equal. Then n_{jk} and n_{kj} are both estimates of the same quantity, νjk, and x_{ih} and x_{hi} similarly estimate X_{ih}. The estimates are combined by averaging, the quantity of interest would then be the degree to which the n_{jk} and x_{ih} exceed these estimates, replacing Equations 21 and 22 with

$$(28) \qquad v_m' = \sum_j \sum_{k>j} \left[n_{jk} - \tfrac{1}{2}(n_{jk} + n_{kj}) \right] , \text{ and}$$

$$(29) \qquad g_m' = \sum_i \sum_{h>i} \left[x_{ih} - \tfrac{1}{2}(x_{ih} + x_{hi}) \right]$$

When the sums are distributed, this results in

$$(30) \qquad v_m' = v_m - \tfrac{1}{2}v ;$$

$$(31) \qquad g_m' = g_m - \tfrac{1}{2}g .$$

The maxima for these two quantities would be $\frac{1}{2}v$ and $\frac{1}{2}g$, respectively. The use of these latter quantities as denominators leads to the following indices of consistency:

$$c_t = \frac{2v_m}{v} - 1 \qquad (32)$$

is an index of consistency of item ordering for a test, and

$$c_p = \frac{2g_m}{g} - 1 \qquad (33)$$

is an index of consistency of person ordering for a test.

Application to Incomplete Data

All of the calculations necessary for c_i and c_p can be carried out for incomplete as well as complete data, but the definitional equations 15, 16, 21, and 22 must be used instead of the simplifications. The latter, as noted later, can be shown to have relationships to more traditional psychometric indices and also to the H measures of Loevinger (1947, 1948).

There may be some concern over how the order is defined so that the above-diagonal sums are as large as possible. For complete data, ordering so that $i < h$ implies $x_i > x_h$ and $j < k$ implies $n_j > n_k$ will lead to larger values for g_m and v_m than any alternative. To prove this, let v_{mt} be the same as (21) except that items j and k are interchanged in the order. This means that certain elements which were above the diagonal have been replaced by their symmetric counterparts: Assuming $j < k$,

$$v_{mt} = v_m + \sum_{a=j}^{k-1} n_{ka} - \sum_{a=j}^{k-1} n_{ak} + \sum_{b=j}^{k-1} n_{bj} - \sum_{b=j}^{k-1} n_{jb} \qquad (34)$$

Then

$$v_{mt} = v_m + \sum_{a=j}^{k-1} (n_{ka} - n_{ak}) + \sum_{b=j}^{k-1} (n_{bj} - n_{jb}) \qquad (35)$$

But for complete data

$$(36) \qquad n_{ka} - n_{ak} = n_k - n_a ,$$

so

$$(37) \qquad v_{mt} = v_m + (k - j)n_k - \sum_{a=j}^{k-1} n_a - (k - j)n_j + \sum_{b=j}^{k-1} n_b ,$$

The summed quantities are identical, so $v_{mt} - v_m$ will be positive if $n_k - n_j$ is. Thus the maximum g_m and v_m are attained when items are in order of difficulty and persons are in order of score, respectively.

The procedure is not as clear-cut when the data are incomplete, but the formulation used in the proof can be used to define a maximizing order. Given the items in an arbitrary order, begin by calculating v_m for that order. For $j = 1$ and $k = x$ calculate $v_m - v_{mt}$ using the indicated sums. If the difference is negative permute j and k. Then repeat the process for all combinations of j and k for which $k - j = x - 2$, permuting those items with negative differences. Then repeat with $k - j = x - 3$, and so on. The "outside-in" character of the process would appear to guarantee that the final value for v_m is the largest possible one, but no proof of this conclusion is offered here.

The v_m index is calculated on the order resulting from the above procedure and is recommended as an overall index of item consistency. The g_m index for person consistency would be calculated in a parallel way and used to measure the even more important property of consistency in ordering of persons.

Completeness of Ordering

Basic concepts

The completeness of the ordering process is only slightly less important than its consistency, particularly as far as persons are concerned. After all, if persons are merely divided into two classes, one consisting of a single individual, then the fact that this is done with perfect consistency is of only minor comfort. Indices of completeness are thus important adjuncts to indices of consistency. Such indices are considered in this section.

There are a number of ways in which completeness can be defined, but the most obvious index is the proportion of pairs for which $x_{ih} - x_{hi} \neq 0$. A more general index is the proportion of pairs for which $|x_{ih} - x_{hi}| > z_{ih2}$, where the latter is a criterion or threshold value as described in the earlier section on notation. That is, we define indices of completeness of person and item ordering as r_p and r_t, respectively, where

(38) $$r_p = \frac{\Sigma\Sigma x^*_{ih2}}{\frac{1}{2}n(n - 1)},$$

and

(39) $$r_t = \frac{\underset{jk}{\Sigma\Sigma} n^*_{jk2}}{\frac{1}{2}x(x - 1)}$$

For complete data and criterion values of zero, this is simply the number of pairs of scores and difficulties that are not tied. In that case, the two indices must influence each other. If there are y different difficulty levels and the data are perfectly consistent, then the largest possible number of different score levels is y + 1. If y + 1 < n, let

(40) $$b = \frac{n}{y + 1}$$

and let I_b be the largest integer $< b$. Then the largest possible value for r_p occurs when some I_b individuals are tied at each of $x(b - I_b)$ scores and $I_b + 1$ are tied at each of the $x(I + 1 - b)$ remaining ones. This can be used to derive an upper bound for the number of different possible scores given the item difficulties. This can be used as a denominator of (38) in place of $\frac{1}{2}n(n - 1)$. A corresponding adjustment can be made in the denominator of (39) to adjust v_t for the scores.

Extended Implied Orders

The preceding developments and proposals have been presented as ones which would apply to dominance matrices which were derived from either complete score matrices or incomplete ones, but in either case from those item-person relations which were actually present in the data. Recently it has been suggested (Cliff, 1975) that the incomplete score matrices employed in interactive testing can be raised to higher powers than two. These higher powers can be used to complete the dominance matrices N and X when the score matrices are so sparse that they do not do so themselves. The process can also be used to compute the odd powers of A which will result in completion of the score matrices themselves. The indices proposed here can be applied to dominance matrices which are derived from such implied score matrices as well as from N and X themselves.

There are several different ways one could proceed in completing dominance matrices. The simplest is to calculate $\sum_p N^p$ and $\sum_p X^p$ for some value of p and then calculate the indices from these matrices. But in effect this is as if the score matrices were other than binary because, e.g.,

$$(41) \qquad N^3 = (\tilde{S}'S\tilde{S}')(S\tilde{S}'S),$$

and the two factors on the right can be thought of as implied score matrices, and they need not be binary.

A more consistent approach is to calculate <u>binary</u> implied score matrices. By the notation suggested earlier, for example, A^{3*} is such a binary matrix whose non-zero elements contain items which are right or wrong by implication as well as directly. Corresponding N and X matrices can then be computed and the indices of consistency and completeness calculated from them. Similarly, A^{5*}, A^{7*},...could be derived using more remote implications, and the corresponding N and X matrices used to give indices for them as well. Thus the indices would be used for implied score matrices as well as observed ones.

The question of how the values in the criterion matrix Z are arrived at is not dealt with here. This is the major statistical inference problem in tailored testing, how it is decided that a response to an item is predictable on the basis of other responses. Different methods of doing that lead to different implied score matrices and the different implied score matrices will have different degrees of consistency. Until more detailed analysis shows otherwise, it must be borne in mind that some methods might lead to spurious degrees of apparent consistency in the ordering. One of the desirable characteristics of a tailored testing method would be that it not lead to such spuriousness, but the problem of how to ensure that seems a difficult one.

Ordinal Information and Its Relation to Traditional Psychometrics

Variance and Dominance

In the complete case, the traditional psychometric indices of location, dispersion, and internal consistency can be related to dominance measures; particularly those related to items. This is to be expected in view of the fact that then the complement matrix $\bar{S}$ can be derived from S by numerical means (Equation 3). Using equations (11) and (19)

$$(42) \qquad v = \Sigma x_i(x - x_i)$$

Then since $\Sigma x_i = n\bar{x}$ and $\Sigma x_i^2 = n\bar{x}^2 + n\sigma_x^2$

$$(43) \qquad v = n\bar{x}(x - \bar{x}) - n\sigma_x^2 ,$$

and

$$(44) \qquad \sigma_x^2 = \bar{x}(x - \bar{x}) - \frac{v}{n}$$

shows the relation between variance and item dominance relations. The relation is complementary rather than direct, which is counterintuitive at first glance.

The maximum possible value for σ_x^2 occurs when the score distribution is dichotomous, and then it is $\bar{x}(x - \bar{x})$. In this case v will equal zero; there will be no item dominance relations. This means all items have exactly the same score profiles; they will be equally difficult but also perfectly consistent. If the items are of equal difficulty but not perfectly consistent then σ_x^2 will not attain this maximum value.

Looked at the other way, $n\bar{x}(x - \bar{x})$ is simply the maximum value v can attain, and it will be denoted v_w.

$$(45) \qquad v_w = n\bar{x}(x - \bar{x})$$

In the present context $\bar{x}(x - \bar{x})$ is taken as a fixed quantity, but it also represents the appropriateness of the items and persons for each other in terms of average difficulty and score levels. Dominance relations occur insofar as items and persons are appropriate to each other; if items are too easy or hard, then $\bar{x}$ or $x - \bar{x}$ will be zero and no dominance will occur. Bearing this as well as the preceding result in mind, variance is proportional to the difference between the number of dominance relations and the maximum possible number that can occur given the average score and difficulty levels.

It may be noted in passing that by this interpretation variance is a pure number, a unitless scalar, since v has persons as units and the constant of proportionality for relating it to variance is n.

Person dominance

It is perhaps curious that total score variance, which is habitually thought of as relating to discrimination among persons, is found to relate to item dominance rather than person dominance relations. The latter, too, are related to a more traditional quantity. Consider (20) using $p_j = n_j/n$:

$$\Sigma n_j(n - n_j) = n^2 \Sigma p_j(1 - p_j) \qquad (46)$$

Thus the number of person dominance relations is proportional to the sum of the item variances. Krus (1974) and Loevinger (1947) note this as the relation between $p_j(1 - p_j)$ and the number of discriminations made by the item.

The Kuder-Richardson formulas

It is of interest to interpret the Kuder-Richardson formulas in terms of dominance. Suppose two items are independent. Then the expected value of n_{jk} is

(47) $$E(n_{jk}) = \frac{1}{n} n_j(n - n_k)$$

This can be used to define v_c, an expectation for v corresponding to the situation of all items being independent.

(48) $$v_c = \sum_j \sum_{k \neq j} E(n_{jk}) ,$$

or for complete data

(49) $$v_c = \frac{1}{n} \sum_{jk}\sum n_j \tilde{n}_k - \frac{1}{n} \sum_j n_j \tilde{n}_j .$$

Substitution and simplification leads to

(50) $$v_c = n\bar{x}(x - \bar{x}) - n\Sigma p_j(1 - p_j)$$

Using our previous expression (43) for v in terms of variance leads to the difference

(51) $$v_c - v = n\sigma_x^2 - n\Sigma p_j(1 - p_j)$$

In present notation, KR20 is

(52) $$\text{KR20} = \frac{x}{x - 1} \left(\frac{\sigma_x^2 - 2p_j(1 - p_j)}{\sigma_x^2}\right)$$

If numerator and denominator are multiplied by n, (45) and (50) lead directly to

(53) $$\text{KR20} = \frac{x}{x - 1} \frac{v_c - v}{v_w - v} .$$

Thus, aside from the correction $x/(x - 1)$, <u>KR20 is the ratio of the difference between the number of dominance relations and the number expected by chance to the difference between v and the maximum possible number</u>. The correction has the effect of making the ratio unity when v takes on its minimum possible value. The foregoing provides still another interpretation of KR20 to go with the recent one provided by Kaiser and Michael (1974).

A second variety of chance expectation for n_{jk} merits consideration. Suppose all items were not only independent but of equal difficulty. That is, there were no true dominance relations at all, the observed ones representing simply random events. Then

$$(54) \qquad E'(n_{jk}) = np(1 - p)$$

for all j and k.

The obvious estimate of p is $\bar{n}/n$, so under these assumptions

$$(55) \qquad E'(n_{jk}) = \bar{n}(n - \bar{n})/n$$

Then we can define a second chance expectation for v:

$$(56) \qquad v_c' = \sum_j \sum_{k \neq j} \bar{n}(n - \bar{n})/n ,$$

which is obviously

$$(57) \qquad v_c' = x(x - 1)\bar{n}(n - \bar{n})/n .$$

Since $x\bar{n} = n\bar{x}$ here

$$(58) \qquad v_c' = n\bar{x}(x - \bar{x}) - n\bar{x}(x - \bar{x})/x$$

or

$$(59) \qquad v_c' = (1 - \frac{1}{x})v_w$$

Then

$$(60) \qquad v_c' - v = n\sigma_x^2 - n\bar{x}(x - \bar{x})/x .$$

Recalling KR21 in present notation leads to

$$(61) \qquad KR21 = \frac{x}{x - 1}\,\frac{v_c' - v}{v_w - v} .$$

Thus KR21 is the ratio of the difference between the number of observed dominances and the number expected under independence and equal difficulty to the difference between the number of observed dominances and the maximum possible number.

The fact that the two principal KR formulas can be expressed in terms of consistency of item dominance, whereas consistency of person is presumably the goal of testing leads to an intriguing notion. This is that the KR formulas should be replaced, or at least supplemented, by ones which express consistency of person dominance. Denoting these as KR_p20 and KR_p21, the duality of items and persons leads immediately to

(62) $$KR_p20 = \frac{n}{n-1}\,\frac{g_c - g}{g_w - g} \,,$$

and

(63) $$KR_p21 = \frac{n}{n-1}\,\frac{g_c' - g}{g_w' - g} \,.$$

Here,

(64) $$g_c = \sum_i \sum_{h \neq i} E(x_{ih})$$

in which

(65) $$E(x_{ih}) = \frac{x_i(x - x_h)}{x}$$

and g_c', and g_w are defined in ways parallel to v_c', and v_w, respectively. In terms of conventional psychometric quantities (σ_n^2 being the variance of the n_j) these are expressed as

(66) $$KR_p20 = \frac{n}{n-1}\left[1 - \frac{\Sigma n_j(n - n_j)}{n^2\sigma_n^2}\right] ,$$

and

$$(67) \qquad KR_p21 = \frac{n}{n-1}\left[1 - \frac{\overline{n}(n-\overline{n})}{n\sigma_n^2}\right]$$

It is conceivable that formulas of the KR type could be generalized to the tailored case, but no way of doing this has presented itself. The difficulty is in defining an analogue of the variance which is strictly comparable in complete and incomplete cases. The incompleteness of the data gives a flexibility which makes it impossible to define v_w, the maximum possible value of v, in other than an arbitrary manner. The concept of a total score is equally slippery with incomplete data. The c indices, on the other hand, depend only on there being orders for items or persons. They are calcuable on any order, even an arbitrary one. Moreover, the procedure outlined above for finding an optimum order seem quite straightforward. Hence the c indices are recommended as appropriate indices for assessing interitem consistency in both complete and incomplete cases.

Loevinger's indices

Loevinger (1947) defines a homogeneity index H_t which depends on V_x, the variance of the present test; V_{hom}, the variance of a perfectly consistent test; and V_{het}, the variance of a test composed of independent items. The latter are analagous in concept to v_m and v_c, respectively. The index turns out to be mathematically related to them.

The index H_t is defined by Loevinger (1947, p. 31) as

$$(68) \qquad H_t = \frac{V_x - V_{het}}{V_{hom} - V_{het}} .$$

In present notation,

(69) $$V_{het} = \frac{1}{n^2} \Sigma n_j(n - n_j) \quad ,$$

and

(70) $$V_{hom} = \frac{1}{n^2} \left[n\Sigma j n_j - n\Sigma n_j - (\Sigma n_j)^2 \right]$$

It is easily seen from the discussion of the KR formulas that

(71) $$n(V_x - V_{het}) = v_c - v \; .$$

Furthermore, from (26) and (50)

(72) $$n(V_{hom} - V_{het}) = v_c - v_m \; .$$

Therefore, Loevinger's index of homogeneity is

(73) $$H_t = \frac{v_c - v}{v_c - v_m} \; .$$

This immediately suggests an index of person consistency defined in terms of the g indices:

(74) $$H_p = \frac{g_c - g}{g_c - g_m} \; .$$

The resemblance of Formula (73) to the KR20 formula coupled with the present concern with dominance relations suggests a further alternative. If there are no true dominance relations among items, then the sample item dominances represent purely chance events, as in the analysis of the KR21 analogue above (Formula 61). Then an alternative to (68) is

(75) $$H_t{}' = \frac{v_c{}' - v}{v_c{}' - v_m}$$

where $v_c{}'$ is defined as in (56).

Still a third homogeneity index suggests itself, one in which v_w is used as an origin rather than v_c or $v_c{}'$:

$$(76) \qquad H_t{}'' = \frac{v_w - v}{v_w - v_m}$$

There would be corresponding homogeneity indices for persons

$$(77) \qquad H_p{}' = \frac{g_c{}' - g}{g_c{}' - g_m}$$

and

$$(78) \qquad H_p{}'' = \frac{g_w - g}{g_w - g_m}$$

The chance expectations v_c and g_c can be generalized to the incomplete case, so H_t and H_p could be used to evaluate the measurement process there. That is, the definition embodied in Equation (48) would still apply, and it could be used to calculate H_t for incomplete data. The individual $E(n_{jk})$ are straightforward to compute but could be time-consuming because they would need to be based on only the persons who took both items. Essentially, n is a variable depending on the j,k pair rather than a constant. In a computerized system the additional requirement would not pose an insuperable difficulty. Some of the simplicity of the c indices would be lost, however, by the introduction of this complication of computing not only v and v_m but v_c as well. Thus the c indices seem still to offer advantages.

The variations $H_t{}'$ and $H_t{}''$, on the other hand, become much too complicated to deal with in the incomplete case because of the difficulty in defining v_w which was noted above in the discussion of the KR formulas.

Item Information

Item Difficulty

The basic information about items which is of concern to the test developer is the item difficulty and discriminating power. These are traditionally measured by proportion passing, which is $1 - p_j$ in present notation, and item-test biserial or point-biserial correlation, respectively. Here we deal with the question of what to use for these purposes in the present context.

The difficulties inherent in measuring difficulty as proportion passing are obvious in the tailored case where the object of the tailoring strategy is to keep the probability of correct response as nearly constant as possible. This fact is the basic motive here although the case of complete data is considered first. We seek indices which not only make intuitive sense in the complete case but also in the tailored one.

One of the theoretical innovations of modern test theory is the placing of persons and items on the same scale. This is true not only of the traceline models (Lord and Novick, 1968), but also of the ordinal model used here. The introductory discussion made note of the fact that it leads to a joint partial order of items and persons. Analagous to the scale value of items on the ability continuum in the traceline models, we have the position of the item in the joint order.

To specify this, it is necessary to know not only the item-person relations but the person-person and item-item ones as well. Here we make use of the relational matrix A^{p*}. In the simplest case, take $A + A^{2*}$. Then we define u_i as follows

$$(79) \qquad u_i = \left[(A + A^{2*})1 - 1'(A + A^{2*}) \right] /(x + n) ,$$

where the index i runs over both items and individuals. The score u_i is thus the difference between an element's "wins" over items and persons and its losses.

For incomplete data, it may be necessary to make use of a "Boolean sum" operator denoted $\ddagger$. Here, $\ddagger a_{ijp}$ is taken in Boolean arithmetic. Then

$$(80) \qquad u_{ip} = \left[(\underset{p}{\ddagger} A^{p*})1 - 1'(\underset{p}{\ddagger} A^{p*}) \right] /(n + x)$$

can be used as a generalized index of the position of the item or person in the order. If p is limited to unity, then u_{ip} is proportional to a right minus wrong score or a fail minus pass index.

There are a couple of additional points that can be noted. First, items and persons are treated together rather than separately or differently. Second, u_{ip} provides an alternative way of defining the appropriate item or person order in order to compute v_m or g_m.

Discrimination Indices

The obvious place to start in defining a "discrimination index" is with components of the v and v_m. The total number of dominances involving item j is

$$(81) \qquad v_j = \sum_{k=1}^{x} (n_{jk} + n_{kj}) .$$

The number which are in the appropriate direction are

$$(82) \qquad v_{mj} = \sum_{k=1}^{j} n_{kj} + \sum_{k=j+1}^{x} n_{jk} .$$

Then $v = \frac{1}{2}\Sigma v_j$ and $v_m + \frac{1}{2}\Sigma v_{mj}$. This immediately suggests that

$$(83) \qquad c_{tj} = \frac{2v_{mj}}{v_j} - 1$$

is an index of consistency for item j.

Complete Data

For a given item, if there is complete data,

$$(84) \qquad v_j = \sum_i \bar{s}_{ij} x_i + \sum_{k \neq j} s_{ij} \bar{x}_i$$

which is

$$(85) \qquad v_j = \sum_i (1 - s_{ij}) x_i + \Sigma s_{ij}(x - x_i).$$

This reduces to the difference between the sum of the total scores of those who got the item right and the sum for those who got it wrong, plus xn_j.

The minimum number of dominations for j, v_{mj}, is the sum of the above diagonal entries for that item:

$$(86) \qquad v_{mj} = \sum_{k=j}^{j-1} n_{kj} + \sum_{k=j+1}^{x} n_{jk} .$$

This can be expressed in terms of the marginals by the same reasoning as for v_m

$$(87) \qquad v_{mj} = \sum_{k=1}^{j-1} n_k - \sum_{k=j+1}^{x} n_k + (n - 2_j + 1)n_j .$$

Thus c_{tj} is computable from item and person totals when the data is complete.

Item indices of the H_t variety can also be used. Recalling Equations

(48) and (49), we define

(88) $$v_{cj} = \sum_{k \neq j} \left[E(n_{jk}) + E(\bar{n}_{kj}) \right] ,$$

and so

(89) $$v_{cj} = \frac{1}{n} \left[\Sigma n_j (n - n_k) + \Sigma n_k (n - n_j) - 2n_j (n - n_j) \right] .$$

Then a homogeneity index for item j can be defined as

(90) $$H_{tj} = \frac{v_{cj} - v_j}{v_{cj} - v_{mj}} .$$

Using the second type of chance expectation used in (54), i.e., that item difficulties are all the same, leads to

(91) $$v_c{}'_j = \frac{2n - 2}{n} \bar{n}(n - \bar{n}) ,$$

and an H'_{tj} may be defined parallel to H_{tj}.

Surprizingly, Loevinger (1947) did not use an H_{tj} like the above to measure item consistency. Rather, she made an interesting departure, in effect measuring the change in g, not v, resulting from deleting the item. This is an interesting approach and has a good deal to be said for it conceptually. It may have been impractical given the computational facilities available at that time to use her approach. On the other hand, Loevinger does use an index similar to (90) in measuring the consistency between a <u>pair</u> of items.

In comparing c_{tj} to H_{tj}, the advantage may lie with the latter since the former will tend to be larger as items' difficulties deviate from .5. For H_{tj}, on the other hand, the bias is, if anything, in the opposite direction.

Parallel indices can be developed for individuals. While persons are normally treated as if they did not vary in their consistency, it is quite conceivable that in fact they do. Certainly in achievement tests the difficulty of the items is largely a function of the manner in which the educational system is organized. Individuals who take an achievement test structured for a particular educational system but who have participated in a different one may well be expected to have low consistency indices. To take account of the possibility of different consistencies for different persons, we define

$$(92) \qquad c_{pi} = \frac{2g_{mi}}{g_i} - 1$$

and

$$(93) \qquad H_{pi} = \frac{g_{ci} - g_i}{g_{ci} - g_{mi}} ,$$

where the g indices are defined parallel to the corresponding v.

Multiple Alternatives

In the foregoing, it has been tacitly assumed that item scoring is dichotomous: Right or wrong, endorse or reject. This limitation is not a necessary aspect of the system, and in many contexts it makes sense to consider multiple alternative responses separately as items in their own right. There are a number of possible different cases; methods of treating a few of the most obvious will be suggested here.

First, there is the multiple choice item. One of the benefits that may be expected to acrue from the computerization of testing is the eventual demise of this type, but the limitations imposed by habit to say nothing of investment, mean that it will be around for some time. The obvious thing to do will be to put in each of the alternatives as a separate item of the score matrix. This raises the question of the score pattern that should be used since we use $\tilde{S}$ as well as S. A logical thing to do is to set $s_{ij} = 1$ and $\tilde{s}_{ji} = 0$ for each correct alternative the individual chooses and for each incorrect alternative he does not choose. On the other hand, $s_{ij} = 0$ and $s_{ji} = 1$ for each correct alternative he fails to choose and each incorrect alternative he does choose.

In the traditional four-choice item where a single alternative must be selected, a person selecting the correct alternative would have $s_{ij} = 1$ for all four possibilities. A person selecting an incorrect alternative would have $s_{ij} = 0$ for the correct alternative and for the incorrect one he chose, but $s_{ij} = 1$ for the two incorrect alternatives he did not select. In this way, "difficulty" and discrimination information could be derived for each alternative.

A second class of multiple alternatives are those where the choices form a logical scale of some kind (e.g., "never," "occasionally," "frequently," "constantly"). In such cases, if it is logical to assume monotonicity for the alternatives, a person who chooses an alternative would be assumed to have also chosen all less extreme alternatives. A conceptually similar case is the timed item, which will be the replacement for the speeded test if the computer ever does take over testing. The time it takes the individual to give the correct response is measured, and time-intervals, e.g. zero to one second, one to two seconds, two to five seconds, five to ten seconds, more than ten seconds, correspond to response alternatives. An individual who responds in three seconds has "passed" the ten-second and the five-second "alternatives" and failed the two-second and one-second ones. Thus the data is treated as if there were an implied logical relation among the alternatives.

These are some possible methods of handling items which take account of multiple alternatives. Others may suggest themselves in the course of time and experience. One thing that needs to be taken account of in evaluating the consistency of items is that the dominance relations among alternatives to the same item are necessarily perfectly consistent. Thus they may inflate the overall consistency estimates. Thus v would have to be corrected for these intra-item dominance relations.

Incomplete Data

The basic definitions of the number of actual, minimum, and chance item dominance relations can be employed for incomplete item data just as they were for the total test. The problems and decisions are just the same.

As noted before in the discussion of v, v_j is calculated from the n_{jk} rather than using the marginals. Similarly, v_{mj} is calculated according to its definition (86) rather than the simplification (87), once an order has been established. Finally, the $E(n_{jk})$ must be calculated on the basis of only those individuals who took both j and k rather than for all. These qualifications mean that the item indices are somewhat more troublesome to compute in the incomplete case, but this is unlikely to cause undue distress since the availability of the computer is assumed.

The item statistics can be applied to the matrices that include implied relations as well as to the directly obtained ones. The caveat that the method of completing the relations matrices may give an artificial degree of apparent consistency should be borne in mind, however.

Discussion

Basic Rationale of the Approach

The developments here followed primarily from considering the wrongs matrix as well as the rights matrix. At first glance, this is a trivial addition to test theory, but it has several advantages.

The first advantage of considering items passing people as well as people passing items is a rather fundamental one. It provides a means of demonstrating that test data can produce an order in the formal sense of a transitive asymmetric relation. If there is a Guttman scale, then $A + A^{(2)}$ is upper triangular when the elements are suitably ordered. Considered as the adjacency matrix of the graph of a relation, it corresponds to the adjacency matrix of an order. Thus the correspondence is very readily and clearly established.

The second conceptual advantage lies in the degree to which it emphasizes the duality between person and item relations. Surely we are not interested in consistency of item ordering per se, but rather in consistency and completeness of person ordering. Item relations are a means to the end of person relations. Use of the complete item-person dominance matrix facilitates the use of the duality in order to develop indices which reflect person relations.

Our primary motivation here has been toward the development of consistency indices which were applicable to tailored tests as well as to standard ones. Obviously, in that case, the wrongs matrix is not completely redundant. There are three possible item-person relations now; right, wrong, and not taken, rather than just two. What we have tried to show here is that even in the incomplete case there are item-item and person-person dominance relations and their consistency and completeness can be evaluated using just the same fundamental quantities as in the complete case.

Thus, a test theory based on the complete matrix of relations among members of an extended set which includes both items and persons has a number of conceptual advantages. Not the least of these is that it provides measures of consistency for tailored tests which are based solely on the test items as they are administered in the tailored context.

Statistical Considerations

The developments here continue the unfortunate psychometric tradition of treating statistical data descriptively rather than inferentially. There is no real justification for this except to say that an inferential system is pointless unless what it is attempting to infer is felt to be worth knowing. Therefore, a set of quantities are presented which are argued to be important descriptors of a measurement system. Unless they are accepted as valid, the additional labor of dealing with the inferential problems that are involved is not justified. To stop at the descriptive level if the descriptors are accepted is equally unjustified.

All that will be done here in considering problems of inference is point out some of the statistical problems. First is the definition of the dimension of generalization. An observed value of one of the consistency indices is based on a sample of items and a sample of persons. In tailored situations, it is furthermore based on a sample of the item-person relations in the sample. Thus in addition to the problems of inferring the value of one of the consistency indices for the whole population of items from a sample of items or for a whole population of persons from a sample of persons, we have that of inferring from a sample of person-item relations to a population of person-item relations. The fact that item-person relations are not sampled randomly in a tailored test but rather are chosen by some scheme which hopes to optimize the selection makes the last inferential problem an especially thorny one. Evaluating the sampling characteristics by Monte Carlo method might be the only feasible method, with conclusions that are consequently sensitive to the design of the sampling experiment.

There is at least one problem of probable bias that should be pointed out. This has to do with v_m and g_m. Items and persons are placed in order on the basis of sample data, so the true order may be different from the sample one. The sample order minimizes v_m and g_m, so some inflation can be expected in the population with consequent shrinkage in consistency indices. A correction for this effect may be feasible, at least in the case of complete data.

It may well be that problems of statistical inference will not be soluble without the adoption of explicit statistical model relating item-person relations to underlying parameters. This has been the route followed by modern test theorists who have followed the traceline or item characteristic curve routes (Birnbaum, 1968; Lord and Novick, 1968; Lazarsfeld and Henry, 1968). This will be unfortunate since the accuracy of the inference will depend on the accuracy of the statistical model, and that will be very difficult to assess.

Relations between Item and Person Measures

In view of the fact that the various v, g,,and c indices are based on the same basic data, closer relations among them might be expected. The only direct relation shown is that inter-relating v, g, v_c, and g_c, Equation (94).

$$g_c - \frac{1}{n} g = v_c - \frac{1}{x} v \tag{94}$$

Further work may well uncover additional ones. This would be a desirable outcome, particularly if it linked item and person indices more closely. That would permit the estimation of one consistency index from the other, and might thus provide support for the use of item consistency to estimate person consistency.

Review and Conclusions

The fundamental premise here is that the set of binary relations between persons and items which is the outcome of the administration of a test is usefully considered as a graph (cf. Harary, Cartwright, and Norman, 1965). The adjacency matrix of such a graph is items-plus-persons by items-plus-persons with the rights matrix S as one part and the wrongs matrix S' as a symmetrically placed second part, as shown in Equation (1)

$$A = \begin{bmatrix} 0 & S' \\ S & 0 \end{bmatrix} . \tag{1}$$

If the score matrix has Guttman form, then a formal correspondence between such a graph matrix and the graph matrix of a partial order can be clearly demonstrated.

The matrix A^2 contains the number of item-item and person-person dominance relations:

$$(4) \qquad A^2 = \begin{matrix} N & 0 \\ 0 & X \end{matrix}$$

Then there are two fundamental measures of dominance

$$(15) \qquad v = \sum_{j}\sum_{k} n_{jk} ,$$

and

$$(16) \qquad g = \sum_{i}\sum_{h} x_{ih} ,$$

where n_{jk} will be the number of persons who get j wrong and k right, and x_{ih} will be the number of items person i gets right which h gets wrong.

If the data are of Guttman form, then $n_{kj} = 0$ for $j < k$ and $x_{hi} = 0$ for $i < h$. In that case, v and g will take on the values v_m and g_m, respectively, where

$$(21) \qquad v_m = \sum_{j} \sum_{k>j} n_{jk} ,$$

and

$$(22) \qquad g_m = \sum_{i} \sum_{h>i} x_{ih} .$$

If there is no dominance at all, then $n_{jk} = n_{kj}$ and $x_{ih} = x_{hi}$ for all pairs of each. This suggests the following indices of consistency of item (c_t) and person (c_p) relations,

(32) $$c_t = \frac{2v_m}{v} - 1$$

(33) $$c_p = \frac{2g_m}{g} - 1$$

A tailored or computerized testing situation is one where it can be true that $s_{ij} = 0$ and $s'_{ij} = 0$, i.e., the person does not take the item. A useful aspect of the foregoing measures and indices is that their definition is equally valid in that case.

Additional special cases of v and g are v_c and g_c and v_w and g_w. The v_c and g_c are expected values of v and g under the assumption of independence:

(48) $$v_c = \sum_j \sum_{k=j} E(n_{jk})$$

(64) $$g_c = \sum_i \sum_{h=i} E(x_{ih})$$

For complete data,

(47) $$E(n_{jk}) = \frac{n_j(n - n_k)}{n}$$

and

(65) $$E(x_{ih}) = \frac{x_i(x - x_h)}{x}$$

where n_j is the number of people who get item j wrong and k right and x_i is the number of items person i gets correct. In the tailored case, the definitions of v_c and g_c stay the same, but n_j, x_i, n and x must be redefined so as to count only persons or items who are in common.

Using v_c and g_c, a pair of consistency indices are

(73) $$H_t = \frac{v_c - v}{v_c - v_m}$$

and

(74) $$H_p = \frac{g - g_c}{g_m - g_c} .$$

H_t is identical to Loevinger's (1947) homogeneity index. Again, they are defined in both complete and incomplete cases.

The quantities v_w and g_w are the maximum values of v and g. These are clearly defined only in the complete case since for incomplete data the definition of a maximum would require elaborate analysis. For complete data, they are

(45) $$v_w = n\bar{x}(x - \bar{x})$$

and g_w is defined similarly.

The familiar KR20 formula can be expressed in these terms

(53) $$KR20 = \frac{x}{x - 1} \frac{v_c - v}{v_w - v}$$

This suggests a person consistency index parallel to KR20,

(62) $$KR_p20 = \frac{n}{n - 1} \frac{g_c - g}{g_w - g} .$$

Thus the KR formulas cannot be generalized to the tailored case unless a means can be found to generalize v_w.

For complete data, there are other simple and important relations

(42) $$v = \Sigma x_i(x - x_i)$$

and

(46) $$g = \Sigma n_j(n - n_j) .$$

Also,

$$(25) \qquad v_m = \sum_j (n - 2j + 1)n_j$$

and

$$(27) \qquad g_m = \sum_i (x - 2i + 1)x_i$$

Completeness of ordering is almost as important as consistency. Thus, two completeness indices were suggested

$$(38) \qquad r_p = \frac{\Sigma\Sigma x^*_{ih}}{\frac{1}{2}n(n - 1)}$$

$$(39) \qquad r_t = \frac{\Sigma\Sigma n^*_{jk}}{\frac{1}{2}x(x - 1)}$$

They are usable in complete as well as incomplete cases.

Item indices of consistency were also proposed. Using only those parts of the sums v, v_m, and v_c which refer to item j, indices parallel to c_t and H_t were derived

$$(83) \qquad c_{tg} = \frac{2vm_j}{v_j} - 1$$

and

$$(90) \qquad H_{tg} = \frac{v_j - v_{cj}}{v_{mj} - v_{cj}}$$

These can be used for the selection, deletion, and evaluation of items in either the complete or incomplete case.

It was also pointed out that the measures could be applied to data where, in the tailored case, some of the elements of the score matrix have been entered by implication rather than directly. This procedure has been suggested by Cliff (1975) as an approach to computer-interactive testing. It was noted, however, that the methods for filling in entries in one of the score matrices may induce an artificial degree of consistency to the data.

Several aspects of the utility of the system proposed here were discussed along with suggested methods of handling certain kinds of data.

References

Airasian, P. W., & Bart, W. M. Ordering theory: A new and useful measurement model. Educational Technology, 1973, 5, 56-60.

Bart, W. M., & Krus, D. J. An ordering-theoretic method to determine hierarachies among items. Educational and Psychological Measurement, 1973, 33, 291-300.

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, Statistical theories of mental test scores. New York: Addison-Wesley, 1968.

Cliff, N. Complete orders from incomplete data: Interactive ordering and tailored testing. Psychological Bulletin, 1975, 82, 289-302.

Coombs, C. H. A theory of data. New York: Wiley, 1964.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Ducamp, A., & Falmagne, J. C. Composite measurement. Journal of Mathematical Psychology, 1969, 6, 359-390.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Guttman, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (Ed.), The prediction of personal adjustment. New York: Social Science Research Council, 1941.

Guttman, L. The basis for scalogram analysis. In S. A. Stouffer (Ed.), Measurement and prediction. New York: Wiley, 1950.

Harary, F., Norman, R., & Cartwright, D. Structural models: An introduction to the theory of directed graphs. New York: Wiley, 1965.

Kaiser, H. F., & Michael, W. B. Domain validity and generalizability. Educational and Psychological Measurement, 1975, 35, 31-35.

Lazarsfeld, P. F., & Henry, N. H. Latent structure analysis. Boston: Houghton Mifflin, 1968.

Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs, 1947, 61 (4, Whole No. 285).

Loevinger, J. The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. Psychological Bulletin, 1948, 45, 507-529.

Reference Note

1. Krus, D. J. Synopsis of Basic theory and techniques of order analysis. (Occasional paper #26) Minneapolis: University of Minnesota, Research, Development, and Demonstration Center in Education of Handicapped Children.

Distribution List

Navy

4 Dr. Marshall J. Farr, Director
Personnel and Training Research Programs
Office of Naval Research (Code 458)
Arlington, VA 22217

1 ONR Branch Office
495 Summer Street
Boston, MA 02210
ATTN: Dr. James Lester

1 ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
ATTN: Dr. Eugene Gloye

1 ONR Branch Office
536 South Clark Street
Chicago, IL 60605
ATTN: Research Psychologist

1 Dr. M.A. Bertin, Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco 96503

1 Office of Naval Research
Area Office
207 West 24th Street
New York, NY 10011

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Dr. H. Wallace Sinaiko
c/o Office of Naval Research
Code 450
Arlington, VA 22217

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Technical Director
Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Assistant Deputy Chief of Naval
Personnel for Retention Analysis
and Coordination (Pers 12)
Room 2403, Arlington Annex
Washington, DC 20370

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Dr. Lee Miller
Naval Air Systems Command
AIR-413E
Washington, DC 20361

1 CDR Paul D. Nelson, MSC, USN
Naval Medical R&D Command (Code 44)
National Naval Medical Center
Bethesda, MD 20014

1 Commanding Officer
Naval Health Research Center
San Diego, CA 92152
ATTN: Library

1 Chairman
Behavioral Science Department
Naval Command & Management Division
U.S. Naval Academy
Annapolis, MD 21402

1 Chief of Naval Education & Training
Naval Air Station
Pensacola, FL 32508
ATTN: CAPT Bruce Stone, USN

1 Mr. Arnold I. Rubinstein
Human Resources Program Manager
Naval Material Command (0344)
Room 1044, Crystal Plaza #5
Washington, DC 20360

1 Dr. Jack R. Borsting
U.S. Naval Postgraduate School
Department of Operations Research
Monterey, CA 93940

1 Director, Navy Occupational Task
Analysis Program (NOTAP)
Navy Personnel Program Support
Activity
Building 1304, Bolling AFB
Washington, DC 20336

1 Office of Civilian Manpower Management
Code 64
Washington, DC 20390
ATTN: Dr. Richard J. Niehaus

1 Office of Civilian Manpower Management
Code 263
Washington, DC 20390

1 Chief of Naval Reserve
Code 3055
New Orleans, LA 70146

1 Assistant to the Assistant Deputy Chief
of Naval Operations (Manpower)
Head, NAMPS, Project Office
Room 1606, Arlington Annex
Washington, DC 20370
ATTN: Dr. Harry M. West

1 Superintendent
Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

1 Mr. George N. Graine
Naval Sea Systems Command
SEA 047C12
Washington, DC 20362

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
ATTN: Dr. Norman J. Kerr

1 Principal Civilian Advisor
for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508
ATTN: Dr. William L. Maloy

1 Director
Training Analysis & Evaluation Group
Code N-00t
Department of the Navy
Orlando, FL 32813
ATTN: Dr. Alfred F. Smode

1 Chief of Naval Training Support
Code N-21
Building 45
Naval Air Station
Pensacola, FL 32508

1 Naval Undersea Center
Code 6032
San Diego, CA 92132
ATTN: W. Gary Thomson

1 LCDR C. F. Logan, USN
F-14 Management System
COMFITAEWWINGPAC
NAS Miramar, CA 92145

1 Navy Personnel Research
and Development Center
Code 01
San Diego, CA 92152

5 Navy Personnel Research
and Development Center
Code 02
San Diego, CA 92152
ATTN: A. A. Sjoholm

2 Navy Personnel Research
and Development Center
Code 310
San Diego, CA 92152
ATTN: Dr. Martin F. Wiskoff

2 Navy Personnel Research
and Development Center
Code 309
San Diego, CA 92152
ATTN: Mr. R. P. Thorpe

1 Navy Personnel Research
and Development Center
San Diego, CA 92152
ATTN: Library

1 Navy Personnel Research
and Development Center
Code 9041
San Diego, CA 92152
ATTN: Dr. J. D. Fletcher

1 D. M. Gragg, CAPT, MC, USN
Head, Educational Programs Development
Department
Naval Health Sciences Education and
Training Command
Bethesda, MD 20014

Army

1 Technical Director
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Headquarters
U.S. Army Administration Center
Personnel Administration Combat Development Activity
ATCP-HRQ
Ft. Benjamin Harrison, IN 46249

1 Armed Forces Staff College
Norfolk, VA 23511
ATTN: Library

1 Commandant
U.S. Army Infantry School
Fort Benning, GA 31905
ATTN: ATSH-DET

1 Deputy Commander
U.S. Army Institute of Administration
Fort Benjamin Harrison, IN 46216
ATTN: EA

1 Dr. Stanley L. Cohen
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Ralph Dusek
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Joseph Ward
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 HQ USAREUR & 7th Army
ODCSOPS
USAREUR Director of GED
APO New York 09403

1 ARI Field Unit - Leavenworth
Post Office Box 3122
Fort Leavenworth, KS 66027

Air Force

1 Research Branch
AF/DPMYAR
Randolph AFB, TX 78148

1 Dr. G. A. Eckstrand (AFHRL/AST)
Wright-Patterson AFB
Ohio 45433

1 AFHRL/DOJN
Stop #63
Lackland AFB, TX 78236

1 Dr. Martin Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Major P. J. DeLeo
Instructional Technology Branch
AF Human Resources Laboratory
Lowry AFB, CO 802[illegible]

1 Dr. Alfred R. Fregly
AFOSR/NL
1400 Wilson Boulevard
Arlington, VA 22209

1 AFHRL/PED
Stop #63
Lackland AFB, TX 78236

Marine Corps

1 Director, Office of Manpower Utilization
Headquarters, Marine Corps (Code MPU)
MCB (Building 2009)
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code RD-1)
Headquarters, U.S. Marine Corps
Washington, DC 20380

1 Chief, Academic Department
Education Center
Marine Corps Development and Education Command
Marine Corps Base
Quantico, VA 22134

1 Mr. E. A. Dover
Manpower Measurement Unit (Code MPI)
Arlington Annex, Room 2413
Arlington, VA 20380

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (G-P-1/62)
U.S. Coast Guard Headquarters
Washington, D.C. 20590

Other DOD

1 Military Assistant for Human Resources
Office of the Secretary of Defense
Room 3D129, Pentagon
Washington, DC 20301

1 Advanced Research Projects Agency
Administrative Services
1400 Wilson Boulevard
Arlington, VA 22209
ATTN: Ardella Holloway

1 Dr. Harold F. O'Neil, Jr.
Advanced Research Projects Agency
Human Resources Research Office
1400 Wilson Boulevard
Arlington, VA 22209

1 Dr. Robert Young
Advanced Research Projects Agency
Human Resources Research Office
1400 Wilson Boulevard
Arlington, VA 22209

1 Mr. Frederick W. Suffa
Chief, Recruiting and Retention Evaluation
Office of the Assistant Secretary of
Defense, M&RA
Room 3D970, Pentagon
Washington, DC 20301

12 Defense Documentation Center
Cameron Station, Building 5
Alexandria, VA 22314
ATTN: TC

Other Government

1 Dr. Lorraine D. Eyde
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. William Gorham, Director
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. Vern Urry
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. Harold T. Yahr
Personnel Research and Development
Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. Andrew R. Molnar
Technological Innovations in
Education Group
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

1 U.S. Civil Service Commission
Federal Office Building
Chicago Regional Staff Division
Regional Psychologist
230 South Dearborn Street
Chicago, IL 60604
ATTN: C. S. Winiewicz

Miscellaneous

1 Dr. Scarvia B. Anderson
Educational Testing Service
17 Executive Park Drive, N.E.
Atlanta, GA 30329

1 Mr. Samuel Ball
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gerald V. Barrett
University of Akron
Department of Psychology
Akron, OH 44325

1 Dr. Bernard M. Bass
University of Rochester
Graduate School of Management
Rochester, NY 14627

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. A. Charnes
BEB 512
University of Texas
Austin, TX 78712

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, NY 14627

1 Dean W. W. Cooper
Carnegie-Mellon University
School of Urban and Public Affairs
Pittsburgh, PA 15213

1 Dr. H. Peter Dachler
University of Maryland
Department of Psychology
College Park, MD 20742

1 Dr. Rene' V. Dawis
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 Dr. Norman R. Dixon
200 South Craig Street
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Marvin D. Dunnette
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 ERIC
Processing and Reference Facility
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Victor Fields
Montgomery College
Department of Psychology
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Visiting Professor
University of California
Graduate School of Administration
Irvine, CA 92664

1 Dr. Robert Glaser, Co-Director
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Mr. Harry H. Harman
Educational Testing Service
Princeton, NJ 08540

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
5640 Nicholson Lane
Rockville, MD 20852

1 Dr. M. D. Havron
Human Sciences Research, Inc.
7710 Old Spring House Road
West Gate Industrial Park
McLean, VA 22101

1 HumRRO Central Division
400 Plaza Building
Pace Boulevard & Fairfield Drive
Pensacola, FL 32505

1 HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921
ATTN: Library

1 HumRRO Central Division/Columbus Office
Suite 23, 2601 Cross Country Drive
Columbus, GA 31906

1 HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921
ATTN: Dr. Robert Vineberg

1 Dr. Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
2001 S Street, N.W., Suite 502
Washington, DC 20009

1 Dr. Milton S. Katz
1500 Park Glen Court
Reston, VA 22090

1 Dr. Robert A. Kaufman
U.S. International University
Graduate School of Human Behavior
Elliott Campus
8655 E. Pomerada Road
San Diego, CA 92124

1 Dr. Steven W. Keele
University of Oregon
Department of Psychology
Eugene, OR 97403

1 Dr. David Klahr
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

1 Dr. Alma E. Lantz
University of Denver
Denver Research Institute
Industrial Economics Division
Denver, CO 80210

1 Mr. W. E. Lassiter
Data Solutions Corporation
6849 Old Dominion Drive, Suite 211
McLean, VA 22101

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Corton Drive
Santa Barbara Research Park
Goleta, CA 93017

1 Dr. William C. Mann
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, CA 90291

1 Mr. Edmond Marks
315 Old Main
Pennsylvania State University
University Park, PA 16802

1 Dr. Leo Munday, Vice President
American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. A. J. Pesch, President
Eclectech Associates, Inc.
P.O. Box 178
North Stonington, CT 06359

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Steven M. Pine
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
University of Southern California
Behavioral Technology Laboratories
3717 South Grand
Los Angeles, CA 90007

1 Dr. Leonard L. Rosenbaum, Chairman
Montgomery College
Department of Psychology
Rockville, MD 20850

1 Dr. George E. Rowland
Rowland and Company, Inc.
P.O. Box 61
Haddonfield, NJ 08033

1 Dr. Benjamin Schneider
University of Maryland
Department of Psychology
College Park, MD 20742

1 Dr. Arthur I. Siegel
Applied Psychological Services
404 East Lancaster Avenue
Wayne, PA 19087

1 Dr. Henry P. Sims, Jr.
Room 630 - Business
Indiana University
Bloomington, IN 47401

1 Dr. C. Harold Stone
1428 Virginia Avenue
Glendale, CA 91202

1 Mr. Dennis J. Sullivan
c/o HAISC, Building 119, M.S. 2
P.O. Box 90515
Los Angeles, California 90009

1 Dr. K. W. Uncapher
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, CA 902[illegible]

1 Dr. David J. Weiss
University of Minnesota
Department of Psychology
N660 Elliott Hall
Minneapolis, MN 55455

1 Mr. George Wheaton
American Institutes for Research
3301 New Mexico Avenue, N.W.
Washington, DC 20016

1 Dr. K. Wescourt
Stanford University
Institute for Mathematical Studies in
the Social Sciences
Stanford, CA 94305

1 Dr. Anita West
Denver Research Institute
University of Denver
Denver, CO 80210

1 Dr. Hanns H. Wolff
8624 Caracas Avenue
Orlando, FL 32807