Journal of Educational Statistics
Fall 1982, *Volume* 7, *Number* 3, *pp.* 215-231

DETECTION OF ABERRANT RESPONSE PATTERNS AND THEIR EFFECT ON DIMENSIONALITY

KIKUMI K. TATSUOKA AND MAURICE M. TATSUOKA
University of Illinois at Urbana-Champaign

Key words: *Error analysis*, *individual consistency index* (*ICI*), *norm conformity index* (*NCI*), *order analysis*

ABSTRACT. Two indices were developed for measuring the degree of conformity or consistency of an individual examinee's response pattern on a set of items. The first, called the norm conformity index (NCI), measures the proximity of the pattern to a baseline pattern in which all 0's precede all 1's when the items are arranged in some prescribed order. The second, called the individual consistency index (ICI), measures the extent to which an individual's response pattern remains invariant when he or she responds to several waves of parallel items. We show how these two indices, used in conjunction with test scores, can spot aberrant response patterns of students requiring detailed error diagnosis and remedial treatment.

The possibility that an examinee obtains correct answers for the wrong reasons when responding to dichotomously scored test items has been largely ignored by psychometricians. Although scattered attempts have been made to give partial credit for partial knowledge, procedures for discrediting correct answers arrived at by incorrect means have typically been confined to the use of formulas for correction for guessing. This lack may not be serious for standardized ability tests, but is very important in the context of achievement testing which is an integral part of the instructional process (Birenbaum & Tatsuoka, in press). Here the test must serve the purpose of diagnosing what type of misconception exists, so that appropriate remedial instruction can be given (Glaser, 1981; Nitko, 1980; Tatsuoka, 1981). This calls for the study of cognitive processes that are used in solving problems, and identifying where the examinee went astray even when the correct answer was produced.

This type of diagnostic testing was pioneered by Brown & Burton (1978). Their celebrated BUGGY is essentially an adaptive, diagnostic testing system, which utilizes network theory for routing examinees through a set of problems in subtraction of positive integers. The branching is such that each successive problem serves to narrow down the scope of "hypotheses" as to the type(s) of misconception(s) held by the examinee until finally a unique diagnosis is made.

Tatsuoka, Birenbaum, Tatsuoka, and Baillie (1980) developed a diagnostic testing system, which differed from BUGGY in that the test was not adaptive but "conventional" (i.e., linear). The test was constructed for use in conjunction with lessons in the addition and subtraction of signed numbers (positive and negative integers) for eighth-grade students and consisted of four parallel subtests of 16 items each. A system of error vectors was developed for diagnosing the type(s) of error committed.

Crucial to this system of error diagnosis is the ability to tell whether and to what extent a response pattern is "typical" or "consistent." We may speak of consistency with respect to either the average response pattern of a group or an individual's own response pattern over time. To measure consistency in these two senses, two related but distinct indices are developed in this paper. They are called the "norm conformity index" (NCI) and "individual consistency index" (ICI), and both are related to Cliff's (1977) group consistency indices.

In particular, it is shown that a certain weighted average of the NCI's of the members of a group yields Cliff's C_{tl}. The higher the value of C_{tl}, the closer the group data is to forming a Guttman scale.

The ICI, on the other hand, measures the degree to which an individual's response pattern remains invariant over time. Thus, for example, in the signed-number test consisting of four parallel subtests, the ICI indicates whether an examinee's response pattern changes markedly from one subtest to the next or remains relatively stable. Low ICI values, indicating instability of response pattern, would suggest that the examinee was still in the early stages of learning, changing his or her method for solving equivalent problems from one wave to the next. A high ICI value, reflecting stability of response pattern, would signal the nearing of mastery or a learning plateau.

Consistency Indices at the Individual Level

Cliff (1977) defined various consistency indices based on the notion of dominance and counterdominance relationships between pairs of items. Some of these are closely related to indices developed in the theory of scalability, originating in the 1930s. Although Cliff's indices are derived from the dominance matrix for the data set of an entire group, they can be expressed as weighted averages of the corresponding indices based on constituent subgroups of examinees (Krus, 1975; Mokken, 1971; Yamamoto & Wise, 1980). What we do in this paper is to carry this process to its logical extreme and define consistency indices at the level of individual responses.

Birenbaum & Tatsuoka (1980, in press) demonstrated that individual response patterns offer useful information for determining any erroneous rule of operation that a given examinee may have used in taking a test. In the context of signed-number arithmetic, Tatsuoka et al. (1980) developed a diagnostic system for identifying erroneous rules by generating "error vectors," each of whose binary elements represents whether or not a specific elementary act (such as adding the absolute values of two numbers) was possibly done. In this paper we develop an index that associates with each response-pattern vector a number between -1 and 1 (inclusive) representing the degree of concordance the vector shows with a Guttman vector of the same length (i.e., with the same number of 1's) when the items are arranged in some purposefully specified order. For instance, they may be arranged—as they are in computing Cliff's indices—in descending order of difficulty for the total group; or they may be arranged in any particular order that suits a given purpose.

Note that group consistency in Cliff's sense is maximized when the items are ordered by difficulty for that group in calculating the indices. Any different ordering would result in a decrease in the value of any of Cliff's consistency indices. The value yielded by each of Cliff's formulas may, therefore, be regarded as a function of item order.

Consider a data set consisting of just one person's response-pattern row-vector **S**. The dominance matrix for this response is

$$\bar{\mathbf{S}}'\mathbf{S} = \mathbf{N} = (n_{ij});\ i,\ j = 1, 2, \ldots, n\ (= \text{number of items}), \tag{1}$$

where $\bar{\mathbf{S}}'$ is the transpose of the complement of **S**. By construction, $n_{ij} = 1$ when the individual gets item i wrong *and* item j right; otherwise $n_{ij} = 0$.

Of course, if the ordering of the items in **S** is changed, the dominance matrix will also change. Consequently, the consistency index associated with response-pattern **S**, defined as

$$C = (2U_a/U) - 1, \tag{2}$$

where $U_a = \Sigma_{i<j} n_{ij}$ (the sum of the above-diagonal elements of **N**), and $U = \Sigma_{i,j} n_{ij}$ (the sum of all the elements of **N**), is a function of the item order, 0. To make this fact explicit, we write $C(0)$. The reader may readily verify that this index has the following properties.

Property 1: $-1 \leqslant C(0) \leqslant 1$.

The upper limit, 1, is attained when **S** is a Guttman vector, with all 0's preceding all 1's while $C(0) = -1$ when **S** is a "reversed Guttman vector," with the 1's coming first.

Property 2: If the order of items is reversed in **S**, the absolute value of $C(0)$ remains unchanged, but its sign is reversed.

Since $U = \Sigma_j \Sigma_i n_{ij} = \Sigma_j \Sigma_i (1 - s_i) s_j$, it is invariant with respect to permutations of the elements of **S**. On the other hand, if the order of the elements of **S** is reversed, so that $s_i = s_{n-i+1}$, U_a for the new dominance matrix will become

$$U_a' = \sum_{i<j} n_{ij}' = \sum_{i<j} (1 - s_{n-i+1}) s_{n-j+1},$$

which can be shown to be equal to $U - U_a$. Therefore

$$\begin{aligned} C'(0) &= [2(U - U_a)/U] - 1 \\ &= (-2U_a/U) + 1 = -C(0). \end{aligned}$$

For situations in which we are interested in the overall consistency of two or more response patterns, we may compute a $C(0)$ associated with a data matrix with multiple rows. We first give the definition for the case of two rows.

Definition: The consistency index $C(0)$ associated with a $2 \times n$ data matrix, comprising two response-pattern vectors $\mathbf{S}_1$ and $\mathbf{S}_2$, is defined by (1) just as in the case of a single response vector. The only change is that the dominance matrix $\mathbf{N}$ is now calculated as follows:

If

$$\mathbf{S} = \left[\frac{\mathbf{S}_1}{\mathbf{S}_2}\right]$$

then

$$\mathbf{N} = [\bar{\mathbf{S}}_1' \bar{\mathbf{S}}_2'] \left[\frac{\mathbf{S}_1}{\mathbf{S}_2}\right] = \bar{\mathbf{S}}_1' \mathbf{S}_1 + \bar{\mathbf{S}}_2' \mathbf{S}_2 = \mathbf{N}_1 + \mathbf{N}_2,$$

where $\mathbf{N}_k$ is the dominance matrix associated with response vector $\mathbf{S}_k$ $(k = 1, 2)$.

Property 3: The consistency index $C(0)$ for the $2 \times n$ matrix $\mathbf{S}$ is a weighted average of the $C(0)$'s for its two rows $\mathbf{S}_1$ and $\mathbf{S}_2$ taken separately. If we let

$$U_k = \sum_{i,j} (\mathbf{N}_k)_{ij} \text{ for } k = 1, 2$$

and

$$U_{ka} = \sum_{i<j} (\mathbf{N}_k)_{ij} \text{ for } k = 1, 2,$$

it follows that U and U_a for $\mathbf{S}$ are given by

$$U = U_1 + U_2$$

and

$$U_a = U_{1a} + U_{2a}.$$

Hence,

$$\begin{aligned} C(0) &= \frac{2(U_{1a} + U_{2a})}{U_1 + U_2} - 1 \\ &= \frac{U_1}{U_1 + U_2} \frac{2U_{1a}}{U_1} + \frac{U_2}{U_1 + U_2} \frac{2U_{2a}}{U_2} - 1 \\ &= \frac{U_1}{U_1 + U_2}\left(\frac{2U_{1a}}{U_1} - 1\right) + \frac{U_2}{U_1 + U_2}\left(\frac{2U_{2a}}{U_2} - 1\right), \end{aligned}$$

or

$$C(0) = w_1 C(0)_1 + w_2 C(0)_2. \tag{3}$$

Remark. The two response patterns $\mathbf{S}_1$ and $\mathbf{S}_2$ may be either those of two individuals or of a single individual taking a set of items on two occasions (as in a repeated-measure design) or two parallel sets of items. In the first case the

$C(0)$ associated with the $2 \times n$ data matrix would be an average $C(0)$ for the pair of individuals; in the second, it would be an average over two measurement occasions for one person.

Norm Conformity Index

By extending the property expressed by (3), it follows that the $C(0)$ associated with an $N \times n$ data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{S}_1 \\ \hline \mathbf{S}_2 \\ \hline \vdots \\ \hline \mathbf{S}_N \end{bmatrix}$$

is a weighted average of $C(0)$'s associated with the individual response-pattern vectors $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N$. In particular, when the items are arranged in descending order of difficulty for the group comprising the N individuals, the $C(0)$ associated with $\mathbf{X}$ is one of Cliff's (1977) consistency indices, C_{tl}. This particular ordering of items is a special case of a broader class of orderings for which we give the name "norm conformity index" to the $C(0)$'s associated with the individual response patterns.

Definition: When the item ordering is in descending order of difficulty for a particular group (designated the "norm group"), the consistency index, $C(0)$, associated with the individual's response pattern, $\mathbf{S}$, is called the norm conformity index, denoted by NCI. Thus, NCI indicates the extent to which a response vector, $\mathbf{S}$, approximates the Guttman vector (in which all the 0's are to the left of the 1's) with the same number of 1's, when the items are arranged in descending order of difficulty for the norm group.

With this definition, plus an expanded version of Property 3, we state the relationship between Cliff's consistency index, C_{tl}, and the NCI's for the individuals in the group as

Property 4: C_{tl} is a weighted average of the $\mathrm{NCI}_k (k = 1, 2, \ldots, N)$, with weights $w_k = U_k/U$; that is,

$$C_{tl} = \sum_{k=1}^{N} (U_k/U)\mathrm{NCI}_k, \tag{4}$$

where $U_k = \Sigma_{i,j}(\mathbf{N}_k)_{ij}$ and $U = \Sigma_{k=1}^{N} U_k$.

Example 1: Let $\mathbf{S}_1 = (01011)$, $\mathbf{S}_2 = (10011)$, and $\mathbf{S}_3 = (00001)$ be the response-pattern vectors for three individuals. Then, the dominance matrix $\mathbf{N}_1$ for the first individual is

$$\mathbf{N}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} [0\ 1\ 0\ 1\ 1] = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

from which we get

$$U_{1a} = 5, \qquad U_1 = 6, \qquad \mathrm{NCI}_1 = 2/3.$$

Similarly, by constructing $\mathbf{N}_2$ and $\mathbf{N}_3$, we obtain

$$U_{2a} = 4, \qquad U_2 = 6, \qquad \mathrm{NCI}_2 = 1/3$$

$$U_{3a} = 4, \qquad U_3 = 4, \qquad \mathrm{NCI}_3 = 1.$$

Hence,

$$\begin{aligned} w_1\mathrm{NCI}_1 + w_2\mathrm{NCI}_2 + w_3\mathrm{NCI}_3 &= (6/16)(2/3) + (6/16)(1/3) + (4/16)(1) \\ &= 5/8. \end{aligned}$$

On the other hand, with

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\overline{\mathbf{X}}'\mathbf{X} = \mathbf{N} = \begin{bmatrix} 0 & 1 & 0 & 1 & 2 \\ 1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$U_a = 13$, $U = 16$, $\mathrm{NCI} = 26/16 - 1 = 5/8$, thus illustrating Property 4.

In the paragraph preceding Property 4, the order of the items was taken to be the order of difficulty for the group of which the individual was a member, for $C(0)$ to be called NCI. Actually, as evident in the formal definition of NCI, the group need not be one to which the individual belongs. It can be any group which the researcher chooses for defining the baseline or "criterion order" of the items; hence, our referring to it as the norm group, and the index as the norm conformity index. Thus, for example, we might be concerned with two groups of students with vastly different instructional backgrounds but similar abilities. It is then possible for the difficulties of various skills to be rather different in the two groups. We may use a conceptually (or intuitively) defined ordering of the items as an hypothetical norm group. We could compute NCI's for members of both Group 1 and Group 2 on the basis of this criterion order, and would probably find the mean NCI for the two groups to be significantly different. The following examples, based on real data, illustrate this.

Example 2: The seventh-grade students of a junior high school were divided at random into two groups, which were given different lessons teaching signed-number operations (Birenbaum & Tatsuoka, in press). One sequence of lessons taught the operations by the Postman Stories approach (Davis, 1964), while the other used the number-line method.

After addition problems had been taught, a 52-item test including both addition and subtraction problems was administered to all students.

A t-test showed no significant difference between the mean test score of the two groups, as indicated in Table I. However, when NCI's were computed for all students, using the item-difficulty order based on a conceptual analysis using an entirely different group, there was a significant difference between the mean NCI of the two groups.

Example 3: Tatsuoka & Birenbaum (1979) demonstrated that proactive inhibition affected the performance on tests in material learned through subsequent instructions. The response patterns of students who studied new lessons written by using a conceptual framework different from that of their previous instructions showed a significantly different performance pattern. By a cluster analysis, four groups among which response patterns are conspicuously different were identified. The NCI values for 91 students based on the order of tasks determined by the proportion correct in the total sample were calculated. The means and standard deviations of the NCI values for the four instructional background groups are shown in Table II.

TABLE I

The Means of Test Scores and the NCI

	Group 1 (N = 67)	Group 2 (N = 62)	
Total Score			
mean	20.06	18.36	$t = 1.190$
SD	8.30	7.88	$p > .05$
NCI			
mean	.375	.216	$t = 3.037$
SD	.230	.292	$p = .003$

TABLE II

Means and Standard Deviations of NCI for Four Groups With Different Instructional Backgrounds

Group	N	Mean	Standard Deviation
1	34	.18	.26
2	27	.41	.28
3	20	.35	.31
4	10	.18	.43

Alternative Definition of C(0)

Up to this point, the U_a and U in (2) defining $C(0)$—and hence NCI as a special case—were defined in terms of the numbers of dominances and counter-dominances between item pairs in the dominance matrix **N**. We now show that U_a can be explicitly defined in terms of the proximity of a response vector **S** to a Guttman vector with the same number of 1's.

Property 5: Let **S** be a response-pattern vector of an examinee on an n-item test, $\mathbf{N} = \bar{\mathbf{S}}'\mathbf{S}$ the associated dominance matrix, and

$$U_a = \sum_{i<j} n_{ij}.$$

Then U_a is also the number of transpositions required to get from **S** to the *reversed* Guttman vector (all 1's preceding the 0's).

Since $n_{ij} = (1 - s_i)s_j$, it follows that

$$U_a = \sum_{i<j} (1 - s_i)s_j$$

is the number of ordered pairs (s_i, s_j) $[i < j]$ of elements of S such that $s_i = 0$ and $s_j = 1$. That is, if for each $s_i = 0$, we count the number of $s_j = 1$ to its right in **S**, then the sum of these numbers over the set of 0's in **S** is equal to U_a. But this is the same as the number of transpositions (interchanges of elements in *adjacent* (0, 1) pairs) needed to transform **S**, step by step, into (1 1...1 0 0 0...0). Thus, NCI is a measure of *remoteness* of **S** from the *reversed* Guttman vector, which is equivalently its *proximity* to the Guttman vector.

Example 4: Let **S** = (01011). Then, **S** can be transformed into (11100) by five successive transpositions:

$$\left(\underline{0\ 1}\ 0\ 1\ 1\right) \rightarrow \left(1\ 0\ \underline{0\ 1}\ 1\right) \rightarrow \left(1\ \underline{0\ 1}\ 0\ 1\right) \rightarrow$$
$$\left(1\ 1\ 0\ \underline{0\ 1}\right) \rightarrow \left(1\ 1\ \underline{0\ 1}\ 0\right) \rightarrow (1\ 1\ 1\ 0\ 0),$$

thus $U_a = 5$ by the present definition. On the other hand,

$$\mathbf{N} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} [0\ 1\ 0\ 1\ 1] = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $U_a = \Sigma_{i<j} n_{ij} = 5$ by the earlier definition.

Note also that, if we denote the number of 1's in the lower triangle of $\bar{\mathbf{S}}'\mathbf{S}$ by U_b, that is,

$$U_b = \sum_{j<i} n_{ij},$$

then U_b is the number of ordered pairs (s_j, s_i) $[j < i]$ of elements of S such that $s_i = 0$ and $s_j = 1$. Hence,

$$U = \sum_{i,j} n_{ij} = U_a + U_b$$

is the number of pairs (s_i, s_j) with $s_i \neq s_j$ that can be formed from the elements of **S**. Thus, $U = x(n - x)$, where x is the number of 1's in **S**, or the test score earned by a person with response pattern **S**. Consequently, U_a/U and U_b/U are the proportions of (0, 1) pairs and (1, 0) pairs, respectively, among all possible ordered pairs (s_i, s_j) $[i > j]$ of unlike elements. When **S** is a Guttman vector $(0\ 0 \ldots 0\ 1\ 1 \ldots 1)$, $U_a = U$ and $U_b = 0$ because all ordered pairs of unlike elements are (0, 1) pairs. Conversely, when **S** is a reversed Guttman vector $(1\ 1 \ldots 1\ 0\ 0 \ldots 0)$, $U_a = 0$ and $U_b = U$. Hence U_a/U ranges from 0 to 1 as an increasing function of the degree to which **S** resembles (or is proximal to) a Guttman vector. Similarly, U_b/U measures the proximity of **S** to a reverse Guttman vector, or its remoteness from a Guttman vector. In fact U_b/U was denoted by U' and proposed as an index of "deviance" of score patterns by van der Flier (1977).

With the above redefinition of U_a and U, the sense in which NCI is a measure of the extent to which a response pattern approximates a Guttman vector should be clear, for

$$\text{NCI} = (2U_a/U) - 1$$

is a rescaling of U_a/U to have limits 1 and −1 instead of 1 and 0.

The concept and properties of transpositions utilized above are well known in the theory of permutation groups (see, e.g., Wielandt, 1964). They have been introduced to the statitsical literature by Kendall (1948/1970) and by Goodman and Kruskal (1954), among others. In fact, our $C(0)$ (and, hence, NCI) can be obtained by computing the Goodman-Kruskal rank coefficient, gamma, between the item-difficulty order for the norm group and the 0-1 scores of the individual, regarded as tied ranks of item difficulty for the individual.[1]

Property 6: Suppose $\mathbf{S}_1$ and $\mathbf{S}_2$ are two n-item response patterns with the same number, x, of 1's, and that $\mathbf{S}_2$ results from $\mathbf{S}_1$ by applying t successive transpositions. Then

$$C(\mathbf{S}_2) = C(\mathbf{S}_1) \pm 2t/x(n - x),$$

where the + sign is taken when $\mathbf{S}_2$ is closer to a Guttman vector than is $\mathbf{S}_1$ and the − sign when the opposite is true.

From Property 5, the U_a associated with a given response pattern **S** is the number of transpositions necessary for getting from **S** to the reversed Guttman

[1] We are indebted to Lawrence Hubert for discovering this relationship and pointing it out to us.

vector with the same number of 1's. Hence, if t is the number of transpositions it takes to get from $\mathbf{S}_1$ to $\mathbf{S}_2$, it follows that

$$U_{2a} = U_{1a} + t$$

when $\mathbf{S}_2$ is closer than $\mathbf{S}_1$ to the Guttman vector, and

$$U_{2a} = U_{1a} - t$$

when the opposite is true. Consequently,

$$\begin{aligned} C(\mathbf{S}_2) &= \frac{2U_{2a}}{U} - 1 = \frac{2(U_{1a} + t)}{U} - 1 \\ &= \frac{2U_{1a}}{U} - 1 + \frac{2t}{U} \\ &= C(\mathbf{S}_1) + \frac{2t}{x(n - x)} \end{aligned}$$

when $\mathbf{S}_2$ is closer than $\mathbf{S}_1$ to the Guttman vector. The sign preceding $2t/x(n - x)$ becomes – when $\mathbf{S}_2$ is farther than $\mathbf{S}_1$ to the Guttman vector.

Example 5: Let $\mathbf{S}_1 = (1\ 0\ 1\ 0\ 1\ 1)$ and $\mathbf{S}_2 = (0\ 1\ 0\ 1\ 1\ 1)$. It takes two transpositions to get from $\mathbf{S}_1$ to $\mathbf{S}_2$:

$(\underline{1\ 0}\ 1\ 0\ 1\ 1) \rightarrow (0\ 1\ \underline{1\ 0}\ 1\ 1) \rightarrow (0\ 1\ 0\ 1\ 1\ 1)$, and $\mathbf{S}_2$ is closer than $\mathbf{S}_1$ to the Guttman vector $(0\ 0\ 1\ 1\ 1\ 1)$. Therefore, by Property 6 we should have

$$\begin{aligned} C(\mathbf{S}_2) &= C(\mathbf{S}_1) + (2)(2)/(4)(2) \\ &= C(\mathbf{S}_1) + 1/2. \end{aligned}$$

For the two response patterns, we have

$$U_{1a} = 5 \text{ and } U_{2a} = 7,$$

so that

$$C(\mathbf{S}_1) = (2)(5)/8 - 1 = 1/4$$

and

$$C(\mathbf{S}_2) = (2)(7)/8 - 1 = 3/4,$$

satisfying the above relation.

Property 7: The weights applied to individual NCI's in computing Cliff's consistency index C_{tl} (cf. Property 4) are invariant to changes in the baseline order of items.

This is true because the weights

$$w_p = \frac{U_p}{\Sigma U_p}$$

depend only on $U_p = x_p(n - s_p)$, where x_p is the total score earned by a person, p—that is, on the number of 1's in response pattern $\mathbf{S}_p$, and not on their positions.

It follows that NCI's associated with response patterns yielding scores close to $n/2$ get high weights while those corresponding to extreme scores get low weights. We see that when the number of persons is large, each w_p is a fairly small positive number, while the NCI has a value between 1 and -1. Negative NCI's are an obstruction to having a large group consistency index, C_{tl}.

Individual Consistency Index

In the preceding sections we defined and described various properties of an index that measures the extent to which an individual's response pattern "conforms" to that of a norm group. That index, the NCI, was found to be a linear transform of a deviance index U' proposed earlier by van der Flier (1977). It could also be regarded as a "backward extention" of one of Cliff's (1977) group consistency indices, C_{tl}, down to the individual level. The other index proposed in this paper is, as far as we are aware, completely novel. The following is the kind of situation in which this index—which we call the individual consistency index (ICI)—finds its use.

When a student is in the process of learning—and hence presumably modifying the cognitive processes by which he or she attempts to solve problems—his or her pattern of responses on successive sets of similar items will change considerably from one set to the next. On the other hand, when the student approaches mastery or a "learning plateau," his or her response pattern will probably remain relatively consistent from one set to the next. It is then desirable to have an index that will allow us to measure the extent to which an individual's response pattern remains unchanged or "consistent" over time. This is what the ICI is designed to do. For ease of exposition, we describe this index by making reference to an actual experimental study.

A 64-item, signed-number test was administered to 153 seventh-grade students at a junior high school. The test consisted of 16 different tasks being tested by four parallel items each. The items were arranged so that four parallel subtests were successively given to each examinee. Within each 16-item subtest, the order of items was randomized. Thus, for each examinee there are four response-pattern vectors with 16 elements each. The ICI is defined on these four replications. To simplify the calculations, let us pretend that the four parallel subtests had only seven items each. Consider a person whose response patterns were as shown in the second column of Table III. Also shown in this table are $U = x(7 - x)$ for each response pattern, the number U_a of transpositions needed to transform each response pattern into a reverse Guttman vector, the $C(0)$ for each response pattern, and the weight to be applied to each $C(0)$ for getting an overall index.

Note that the weighted average

$$\sum_{j=1}^{4} w_j C(0)_j = -.143$$

would be Cliff's consistency index, C_{tl}, *if* the four response patterns of Table III were those of four persons instead of one, *and* if the items had been arranged in their difficulty order for the group of four. But, of course, neither of these conditions is satisfied here, so the value $-.143$ just calculated is meaningless.

Let us now rearrange the items (or rather the sets of parallel items) in their order of difficulty for the person under consideration, which is (2, 4, 5, 7, 3, 1, 6). The first four-item sets (2, 4, 5, 7) are tied in difficulty, each having zero correct responses. The order in which these sets are arranged among themselves is immaterial; all orders yield the same result. The response patterns and other quantities occurring in Table III now become as shown in Table IV, which also has a new column showing the number of transpositions, t_j, necessary to get from the jth response pattern in Table III to the new one here. Note that the new $C(0)$ for each response pattern satisfies Property 6:

$$C(0')_j = C(0)_j + 2t_j/U_j.$$

The weighted average of the new $C(0)$ values is

$$\sum_{j=1}^{4} w_j C(0')_j = .952.$$

This is what we call ICI.

TABLE III

Four Response Patterns and Various Quantities Associated with Them

Parallel Test #(j)	Response Pattern	U_j	U_{ja}	$C(0)_j$	w_j
1	(0010010)	10	6	.200	.238
2	(1000010)	10	4	−.200	.238
3	(1000010)	10	4	−.200	.238
4	(1010010)	12	4	−.333	.286

TABLE IV

Response Patterns Resulting from those in Table III by Arranging the Items in Difficulty Order, and Various Associated Quantities

Parallel Test #(j)	Response Pattern	t_j	U_j	U'_{ja}	$C_p(0')_j$	w_j
1	(0000101)	3	10	9	.8	.238
2	(0000011)	6	10	10	1.0	.238
3	(0000011)	6	10	10	1.0	.238
4	(0000111)	8	12	12	1.0	.238

Definition: Given a set of response patterns shown by a single individual on a set of parallel tests, we arrange the parallel items in their overall order of difficulty for the individual and compute the $C(0)$ for each response pattern thus modified. If we now form a weighted average of these $C(0)$'s *as though* we were computing Cliff's C_{tl} in accordance with Property 4, the result is the ICI.

Remark: Note that ICI is an attribute of a single individual, not of a group as is Cliff's consistency index. Also, ICI differs from NCI in that the latter (also an individual attribute) depends on the baseline order of items (i.e., the difficulty order in some group specified as the norm group), whereas ICI is computed for an individual with no reference to any group. Rather, ICI requires that the individual in question has taken two or more parallel tests, and measures the consistency of his or her response patterns across these parallel tests.

Property 8: Because the parallel items are arranged in their order of difficulty for the individual in question when ICI is computed, while they are arranged in their order of difficulty for a norm group when NCI is computed, it follows that

$$\text{ICI} \geqslant \text{NCI}$$

for each examinee.

Practical Application of ICI

Because ICI measures the extent to which the examinee consistently uses some rule of operation, having a high ICI can be good or bad, depending on whether that rule is the correct one or an incorrect one. Therefore, having a high ICI and a high test score is the best condition to be in, while a high ICI combined with a low test score is the worst possible state of affairs—both from the examinee's standpoint *and* from the perspective of the researcher who would use item response theory (IRT). For the student, the prognosis is bad because such a combination means that he or she is "hooked" on an erroneous rule of operation, which suggests that he or she has some misconception that may be difficult to remedy. For the application of IRT, the high-ICI-low-score combination is troublesome because it is the presence of examinees of this category that is most likely to violate any unidimensionality of the test data.

This can be illustrated in conjunction with the study of Birenbaum & Tatsuoka (1980) mentioned earlier. They administered a 64-item, free-response test in signed numbers, consisting of four parallel subtests of 16 items each, to a group of 153 seventh-grade students. A principal components analysis, with the item scored in the usual way (1 or 0 for right or wrong), yielded results that did not begin to approach unidimensionality. Next, Birenbaum and Tatsuoka (1980, in press) modified the data by scoring an item 0 when it was judged to have been correctly answered by use of an erroneous rule. A new principal-components analysis was done using these modified data. The change between the two analyses was dramatic: the dimensionality of the data became much more clear-cut and the item-total biserial correlations became much higher, while the means of the 16 tasks (each represented by four parallel items) did not change drastically.

The act of modifying scores in the above manner would be hard to justify in practice. For the purpose of obtaining nearly unidimensional data sets, it would be less controversial simply to delete cases that appear to be "false corrects." Even so, it would be desirable to have a more objective method for selecting the cases to weed out than the somewhat judgemental one used by Birenbaum and Tatsuoka. ICI, when used along with the total score, provides a more objective screening procedure.

Specifically, a cutoff value of .90 for ICI was determined from error-analysis considerations, and a cutoff score of 52 was similarly decided on.[2] It turned out that, among students who had errors of types that could be unambiguously diagnosed by the error-vector system, about 89 percent had ICI $\geqslant$ .90 and total score less than 52. When these students were removed from the sample, an approach to unidimensionality was found and the iterative estimation of IRT parameters showed convergence. Details of these results are described in Tatsuoka & Tatsuoka (1981).

It should be noted that unidimensionality in the factor-analysis and Guttman-scale senses are two different things. The indices NCI and ICI are measures of scale homogeneity associated with the item-type difficulties based on a group and on an individual, respectively. On the other hand, factor-analytic unidimensionality measures homogeneity of the content of items as reflected by high correlations among item scores. Nevertheless, there is no *inconsistency* in speaking of a group composed of individuals with high ICIs and the data set as a whole being factor analytically unidimensional. The one does not imply the other, but nor do they preclude each other, because a high ICI simply means that a set of responses by one individual comes close to forming a Guttman scale, and says nothing about the whole group's data set approximating such a scale. In fact, a high ICI combined with a low total score is a danger signal, as mentioned before, because it implies that the individual is consistently using some incorrect rule. If there are many examinees of this type in a group, it would certainly be detrimental to the chances of the group data set's being factorially unidimensional, because various different incorrect rules are bound to be used by different individuals leading to spuriously correct answers on different items.

On the other hand, it was revealed by the error diagnostic program SIGNBUG (Tatsuoka & Baillie, 1982) that the subgroup of examinees *excluding*

[2] This is four points more than the maximum score, 48, achievable by consistent use of the "best" erroneous rule.

those with high ICI and low total score consisted predominantly of those who had used various mixes of the right rule and what appear to be random errors (see Tatsuoka & Tatsuoka, 1981). Thus, it is more likely for such a data set to be unidimensional.

Discussion

The two indices for measuring consistency (in two senses) of individual response patterns, NCI and ICI, were originally both developed as potential tools for assisting in the extraction of subsamples of examinees for whom the data are unidimensional or nearly so. Earlier work toward extracting unidimensional subsets (of examinees, items or both) has been done by Sato (1975) among others. The approach taken by Sato and his coworkers utilizes properties of a device known as the S-P (for "Student-Problem") chart (Sato, 1975; see also, Tatsuoka, 1979).

Our approach may be regarded, in a sense, as an attempt to dispense with the "excess baggage" of S-P charts while retaining the order-theoretic foundations on which they rest. It turned out, as we have seen, that the NCI was a sort of backward extension of one of Cliff's (1977) group consistency indices to the individual level. It also turned out that NCI was not very useful for the originally intended purpose of extracting unidimensional subgroups. Rather, it was found to be more useful in highlighting the different response patterns that are typical of individuals with different instructional backgrounds, as illustrated in examples 2 and 3 above, and also in greater detail by Harnisch and Linn (1981).

The other index developed in this paper, ICI, was found to be quite useful for the intended purpose of extracting unidimensional subgroups of examinees —or, more precisely speaking, for identifying individuals who could be removed from a sample to improve the approximation to unidimensionality exhibited by the data matrix of the remaining group. A caveat here is that the number of individuals removed must not be excessively large; otherwise, the item intercorrelations may be seriously attenuated by decrease of variance, and the result would be detrimental to achieving unidimensionality. We mention in passing that, due to the duality between examinees and items (depending on whether one reads data matrices row-wise or columnwise), an analogous index could be defined that would assist in the extraction of unidimensional subgroups of items.

The one drawback of ICI is that it requires a test consisting of at least two (preferably three or four) parallel subtests. In practice, such tests seldom exist. Fortunately, however, it has proved possible to develop several other indices as continuous generalizations of Sato's (1975) caution index, and one of these acts much like ICI *without* the need to have several parallel subtests in the measuring instrument. These "generalized caution indices" are discussed in a recent paper by Tatsuoka & Linn (in press).

Acknowledgement

This research was sponsored by the Personnel and Training Research Program, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-79-C-0752. Contract Authority Identification Number NR 150-415.

References

Birenbaum, M., & Tatsuoka, K. K. *The use of information from wrong responses in measuring students' achievement* (Research Report 80-1 ONR). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1980. (ERIC Document Reproduction Service No. ED 202 915)

Birenbaum, M., & Tatsuoka, K. K. On the dimensionality of achievement test data. *Journal of Educational Measurement*, in press.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 1978, *2*, 155–192.

Cliff, N. A theory of consistency or ordering generalizable to tailored testing. *Psychometrika*, 1977, *42*, 375–399.

Davis, R. B. *The Madison Project, discovery in mathematics; a text for teachers*. Reading, Mass.: Addison-Wesley, 1964.

Glaser, R. *The future of testing: A research agenda for cognitive psychology and psychometrics* (Technical Report). Pittsburgh, Pa.: University Pittsburgh, Spring 1981.

Goodman, L. A., & Krus, W. H. Measures of association for classifications. *Journal of the American Statistical Association*, 1954, *49*, 732–764.

Harnisch, D. L., & Linn, R. L. Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *The Journal of Educational Measurement*, 1981, *18*, 39–87.

Kendall, M. G. *Rank correlation methods* (4th ed.). London: Charles Griffin, 1970. (Originally published, 1948)

Krus, D. J. *Order analysis of binary data matrices*. Los Angeles: Theta Press, 1975.

Mokken, R. J. *A theory and procedure of scale analysis: With applications in political research*. The Hague: Mouton, 1971.

Nitko, A. J. Criterion-referencing schemes. In S. T. Mayo (Ed.), *New directions for testing and measurement: Interpreting test performance* (6). San Francisco: Jossey-Bass, 1980.

Sato, T. *The construction and interpretation of S-P tables*. Tokyo: Meiji Tosho, 1975. (In Japanese)

Tatsuoka, K. K. *An approach to assessing the seriousness of error types and predictability of future performance* (Research Report 81-1 ONR). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, February 1981. (ERIC Document Reproduction Service No. ED 205 538)

Tatsuoka, K. K., & Baillie, R. SIGNBUG: *An error diagnostic computer program for signed-number arithmetic on the PLATO system*. Urbana, Ill.: University of Illinois, Computer-based Research Laboratory, 1982.

Tatusoka, K. K., & Birenbaum, M. *The danger of relying solely on diagnostic adaptive testing when prior and subsequent instructional methods are different* (CERL Report E-5). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, March 1979. (ERIC Document Reproduction Service No. Ed 103 608)

Tatsuoka, K. K, & Birenbaum, M. The effect of different instructional methods on achievement tests. *Journal of Computer-based Instruction*, 1981, *8*(1), 1–8.

Tatsuoka, K. K., Birenbaum, M., Tatsuoka, M. M., & Baillie, R. *A psychometric approach to error analysis on response patterns* (Research Report 80-3 ONR). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1980. (ERIC Document Reproduction Service No. ED 195 577)

Tatsuoka, K. K., & Linn. R. L. Indices for detecting unusual response patterns: Links between two general approaches and potential applications. *Journal of Applied Psychological Measurement*, in press.

Tatsuoka, K. K., & Tatsuoka, M. M. *Spotting erroneous rules of operation by the Individual Consistency Index* (Research Report 81-4 ONR). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1981.

Tatsuoka, M. M. Recent psychometric developments in Japan: Engineers tackle educational measurement problems. *ONR Scientific Bulletin*, 1979, *4*(1), 1–7

van der Flier, H. Environmental factors and deviant response patterns. In Y. H. Poortinga (Ed.), *Basic problems in cross cultural psychology*. Amsterdam: Swets & Seitlinger, B. V., 1977.

Wielandt, H. *Finite permutation groups*. New York: Academic Press, 1964.

Yamamoto, Y., & Wise, S. *Extracting unidimensional chains from multi-dimensional data sets: A graph theory approach* (Research Report 80-2 ONR). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1980.

Authors

TATSUOKA, KIKUMI K. Senior Specialist in Automated Education and Assistant Professor of Educational Psychology, 252 ERL, University of Illionis, 103 S. Mathews, Urbana, IL 61801. *Specializations*: Computerized adaptive testing; psychometrics, especially item response theory.

TATSUOKA, MAURICE M. Professor of Psychology and Educational Psychology, 220 Education Bldg., University of Illinois, 1310 S. Sixth St., Champaign, IL 61820. *Specializations*: Statistics, especially multivariate analysis, in behavioral research; educational measurement.