Indices for Detecting Unusual Patterns: Links Between Two General Approaches and Potential Applications

Kikumi K. Tatsuoka and Robert L. Linn
University of Illinois

Two distinct approaches, one based on item response theory and the other based on observed item responses and standard summary statistics, have been proposed to identify unusual response patterns of responses to test items. A link between these two approaches is provided by showing certain correspondences between Sato's S-P curve theory and item response theory. This link makes possible several extensions of Sato's caution index that take advantage of the results of item response theory. Several such indices are introduced and their use illustrated by application to a set of achievement test data. Two of the newly introduced extended indices were found to be very effective for purposes of identifying persons who consistently use an erroneous rule in attempting to solve signed-number arithmetic problems. The potential importance of this result is briefly discussed.

Several authors have recently shown an interest in using information from patterns of responses to test items to extract information not contained in the total score. A variety of purposes have been envisioned for use of the additional information. Wright (1977), for example, referred to identification of "guessing, sleeping, fumbling, and plodding" (p. 110) from the plots of residual item scores based on the differences between item responses and the expected responses for an individual based on the Rasch model. Levine and Rubin (1979) discussed response patterns that are "so atypical . . . that his or her aptitude test score fails to be a completely appropriate measure" (p. 269). Trabin and Weiss (1979) also investigated measures to identify unusual response patterns. Sato (1975) proposed a "caution" index, which is intended to identify students whose total scores on a test must be treated with caution. Tatsuoka and Tatsuoka (1980) and Harnisch and Linn (1981) have discussed the relationship of response patterns to instructional experiences and the possible use of item response pattern information to help diagnose the types of errors a student is making.

Indices of the degree to which an individual's pattern of responses is unusual are conveniently classified into two general types: (1) those that use item response theory (IRT) to identify unusual patterns and (2) those that rely only on observed item responses and standard summary statistics based on those responses (e.g., the number or proportion of people in a norm group answering an item correctly). The work of Wright (1977; Wright & Stone, 1979), Levine and Rubin (1979) and Trabin and Weiss (1979) are examples of approaches based on IRT, while the work of Sato (1975), Donlon and Fischer (1968), Tatsuoka and Tatsuoka (1980), and Harnisch and Linn (1981) are of the latter type.

The primary purpose of this paper is to develop a link between these two general approaches. More specifically, correspondence will be shown between Sato's (1975) S-P curve theory and test response curves and "group response curves" developed from IRT. Also, Sato's caution index defined in S-P curve theory is generalized into a continuous domain utilizing IRT. That is, S-P curve theory and the caution index were originally developed in a discrete domain of 0–1 scoring, but this study extends the theory to a more general case of probabilities.

First, Sato's S-P curve theory will be reviewed. Next, a group response curve is developed for the one-parameter logistic model. The group response curve is based on the dualistic nature of the one-parameter logistic model, which depends on the choice of fixed and random parameters in the model. Two- and three-parameter logistic models are briefly discussed, with special attention given to problems with person and group response curves in these models. Then, an extended caution index with several special areas which are applicable to IRT will be presented. Finally, a discussion follows about applications of the new caution indices for the detection of anomalous response patterns in the two-parameter model case.

S-P Curve Theory

Sato's (1975) caution index is applicable to either an item or an individual examinee. In either form the index is conveniently obtained from a specially arranged table of binary item scores referred to as an "S-P table." The S-P table, the associated S-P curves and various indices, such as the caution index, are widely used in Japan for diagnosing student performance, for detecting aberrant response patterns, and for assessing the quality of a test or instructional sequence.

The S-P table is a data matrix in which the students (represented by rows) have been arranged in descending order of their total test scores from top to bottom and the items (represented by columns) have been arranged in ascending order of difficulty from left to right. A hypothetical S-P table is shown in Table 1. The solid stair-step line is the student curve (S-curve). For each person, represented by a given row, a vertical line is drawn to the right of the nth cell from the left, where n is the number of correct answers obtained by that person. The S-curve is then obtained by connecting the right edge of the nth cell of each row. The problem curve (P-curve) is drawn in an analogous fashion by counting down from the top the number of cells equal to the number of students who correctly answered the item corresponding to a given column. The P-curve for the data in Table 1 is shown by the dashed line.

Let y_{ij} be the binary response for student (row) i to item (column) j of the S-P table. Row and column sums are denoted by $y_{i.}$ and $y_{.j}$, respectively. The total number of 1's in the S-P table is denoted $y_{..}$ and the proportion of correct responses by $P_{i.}$, $P_{.j}$ and $P_{..}$ for the row, column, and entire table, respectively. As can be seen in Table 1, the S-curve is the step function ogive of the cumulative distribution function of total scores, $y_{i.}$, for the 15 students and the P-curve is the corresponding function of $y_{.j}$, the number of correct answers for the 10 items.

If the S-curve is held invariant and all the 0's to the left of the S-curve are changed to 1's and all the 1's to the right of the same curve to 0's, the result is the S-P table shown in Table 2 and the S-curve is called a perfect S-curve. The entries in Table 2 are denoted M_{ij}^s. Similarly, a perfect P-curve will be obtained, which is shown in Table 3. The entries in the new table are denoted by M_{ij}^p. As can be seen, $M_{i.}^s = y_{i.}$ for all i, which corresponds to the fact that the S-curve is unchanged as the result of changing the cell entries from y_{ij} to M_{ij}^s. The values of the column sums for Tables 1, 2, and 3, i.e., $y_{.j}$ and $M_{.j}^p$ are not in general equal, however.

Sato (1975) defined a caution index for subject i by taking the ratio of two covariances. The numerator of the ratio is the covariance of the observed row vector i, (y_{ij}) $j = 1, \ldots, n$, and the sum-of-column vector, $(y_{.j})$, $j = 1, 2, \ldots, n$; and the denominator is the covariance of the corresponding scores, assuming

Table 1
A Hypothetical Score Matrix (y_{ij}) with

S- Curve (solid line) and P- Curve (dotted line)

Subject i	1	2	3	4	5	6	7	8	9	10	$y_{i.}$	$P_{i.}$	$M_{i.}^S$
1	1	1	1	1	1	1	1	1	1	1	10	1.0	10
2	1	1	1	1	1	1	1	1	1	0	9	0.9	9
3	1	1	1	1	1	0	1	1	0	1	8	0.8	8
4	1	0	1	1	1	1	0	1	0	0	6	0.6	6
5	1	1	1	1	0	1	0	0	1	0	6	0.6	6
6	1	1	1	0	1	0	1	0	1	0	6	0.6	6
7	1	1	1	1	0	0	1	0	0	0	5	0.5	5
8	1	1	1	0	1	1	0	0	0	0	5	0.5	5
9	1	0	0	1	0	1	0	1	1	0	5	0.5	5
10	1	1	0	1	0	0	1	0	0	1	5	0.5	5
11	0	1	1	1	1	0	0	0	0	0	4	0.4	4
12	1	0	0	0	1	1	0	0	0	1	4	0.4	4
13	1	1	0	0	0	1	0	0	0	0	3	0.3	3
14	1	0	1	0	0	0	0	0	0	0	2	0.2	2
15	0	1	0	0	0	0	0	0	0	0	1	0.1	1
$y_{.j}$	13	11	10	9	8	8	6	5	5	4	$y_{..}$ = 79		
$P_{.j}$	.87	.73	.67	.60	.53	.53	.40	.33	.33	.27	$P_{..}$ = .527		
$M_{.j}^P$	13	11	10	9	8	8	6	5	5	4			

the S-curve is perfect, (M_{ij}^s), $j = 1, ..., n$, and the column-sum vector $(y_{.j})$, $j = 1, 2, ..., n$. More specifically, the caution index C_i for subject i is given by

$$C_i = 1 - \frac{\sum\limits_{j=1}^{n}(y_{ij} - P_{i.})(y_{.j} - P_{..})}{\sum\limits_{j=1}^{n}(M_{ij}^S - P_{i.})(y_{.j} - P_{..})} \qquad [1]$$

and the caution index, C_j, for item j is given by

$$C_j = 1 - \frac{\sum\limits_{i=1}^{N}(y_{ij} - P_{.j})(y_{i.} - P_{..})}{\sum\limits_{i=1}^{N}(M_{ij}^P - P_{.j})(y_{i.} - P_{..})} . \qquad [2]$$

The second term of the caution index for item j is the ratio of two covariances: the numerator is the covariance of column vector j, (y_{ij}) and $(y_{i.})$, $i = 1, ..., N$, and the denominator is the covariance of the vectors $(y_{i.})$ and (M_{ij}^P), $i = 1, 2, ..., N$. The value of the denominator is considered as a norm value to standardize the numerator.

Table 2
Perfect S-Curve Obtained by Changing 1's to the Right
of S-Curve to 0 and 0's to the Left to 1

Subject i	Item j										$y_{i.}$	$M^S_{i.}$
	1	2	3	4	5	6	7	8	9	10		
1	1	1	1	1	1	1	1	1	1	1	10	10
2	1	1	1	1	1	1	1	1	1	0	9	9
3	1	1	1	1	1	1	1	1	0	0	8	8
4	1	1	1	1	1	1	0	0	0	0	6	6
5	1	1	1	1	1	1	0	0	0	0	6	6
6	1	1	1	1	1	1	0	0	0	0	6	6
7	1	1	1	1	1	0	0	0	0	0	5	5
8	1	1	1	1	1	0	0	0	0	0	5	5
9	1	1	1	1	1	0	0	0	0	0	5	5
10	1	1	1	1	1	0	0	0	0	0	5	5
11	1	1	1	1	0	0	0	0	0	0	4	4
12	1	1	1	1	0	0	0	0	0	0	4	4
13	1	1	1	0	0	0	0	0	0	0	3	3
14	1	1	0	0	0	0	0	0	0	0	2	2
15	1	0	0	0	0	0	0	0	0	0	1	1
$y_{.j}$	13	11	10	9	8	8	6	5	5	4	79	79
$M^P_{.j}$	15	14	13	12	10	6	3	3	2	1		

It can be said that this ratio in the above caution index is equal to the ratio of the traditional discrimination index, r_j (item-total correlation) to the standardized (or ideal in a sense illustrated in Table 2) discriminating index, r'_j, for item j. That is,

$$\frac{\text{cov}_j(y_{ij}, y_{i.})}{\text{cov}_j(M^P_{ij}, y_{i.})} = \frac{\dfrac{\text{cov}_j(y_{ij}, y_{i.})}{\sigma_j(y_{ij})\,\sigma(y_{i.})}}{\dfrac{\text{cov}_j(M^P_{ij}, y_i)}{\sigma_j(M^P_{ij})\,\sigma(y_i)}} = \frac{r_j}{r'_j}. \qquad [3]$$

It is clear that $\Sigma(y_{ij} - P_{.j})^2 = \Sigma(M^P_{ij} - P_{.j})^2$ because the number of 1's in column j is invariant, as can be seen in Tables 1 and 2, so the number of 1's in the column vector j, (M^P_{ij}) and (y_{ij}) are the same. Therefore, the two variances $\sigma^2_j(y_{ij})$ and $\sigma^2_j(M^P_{ij})$ are equal.

The Extended Caution Index in Conjunction with Item Response Theory

Test and Group Response Curves: One-Parameter Logistic Model

According to the one-parameter logistic model, the item response curve (or item characteristic curve or item response function) may be written

$$P_{b_j}(\theta) = \frac{1}{1+\exp[-D(\theta-b_j)]}, \quad j=1,2,\ldots,n \tag{4}$$

where θ is the latent ability, b_j is the difficulty of item j and D is a constant, which is set equal to 1.7 for convenience of comparison to the normal ogive model (see Lord & Novick, 1968, p. 400). In the above equation, b_j is fixed and θ is a random variable.

   Although in practice, the number of items, n, is a finite number, it is useful to consider b as a continuous variable. By holding θ_i fixed and treating b as a continuous variable, the dual function, $S_{\theta_i}(b)$, of the one-parameter logistic function may be defined,

$$S_{\theta_i}(b) = \frac{1}{1+\exp[-D(\theta_i-b)]}, \quad i=1,2,\ldots,N \ . \tag{5}$$

Of course, the expression

$$\frac{1}{1+\exp[-D(\theta_i-b_j)]} \tag{6}$$

may be considered to be a function of either θ or b. By choice of which variable is fixed, the function may be used to define either the item response curve, $P_{b_j}(\theta)$ or the person response curve $S_{\theta_i}(b)$ (see

Table 3
Perfect P-Curve Obtained by Changing 1's Below
the P-Curve to 0 and 0's Above to 1

Subject i	Item j 1	2	3	4	5	6	7	8	9	10	$y_{i.}$	$M^S_{i.}$
1	1	1	1	1	1	1	1	1	1	1	10	10
2	1	1	1	1	1	1	1	1	1	1	9	10
3	1	1	1	1	1	1	1	1	1	1	8	10
4	1	1	1	1	1	1	1	1	1	1	6	10
5	1	1	1	1	1	1	1	1	1	0	6	9
6	1	1	1	1	1	1	1	0	0	0	6	7
7	1	1	1	1	1	1	0	0	0	0	5	6
8	1	1	1	1	1	1	0	0	0	0	5	6
9	1	1	1	1	0	0	0	0	0	0	5	4
10	1	1	1	0	0	0	0	0	0	0	5	3
11	1	1	0	0	0	0	0	0	0	0	4	2
12	1	0	0	0	0	0	0	0	0	0	4	1
13	1	0	0	0	0	0	0	0	0	0	3	1
14	0	0	0	0	0	0	0	0	0	0	2	0
15	0	0	0	0	0	0	0	0	0	0	1	0
$y_{.j}$	13	11	10	9	8	8	6	5	5	4	79	
$M^P_{.j}$	13	11	10	9	8	8	6	5	5	4	79	

Lumsden, 1978; Trabin & Weiss, 1979). Hence, the variable described within the parentheses of the function is considered as a random variable and the subscript variable is a fixed variable.

The curves for the pair of functions, $P_{b_j}(\theta)$ and $S_{\theta_i}(b)$ are symmetric about the vertical axis at $\theta = \theta_0$ (or, equivalently, $b = b_0$), provided $\theta_0 = b_0$. As illustrated in Figure 1, however, the item response curve (IRC) and the person response curve (PRC), intersect at

$$(\theta_0 + b_0)/2 \text{ if } \theta_0 \neq b_0 \qquad [7]$$

and the test response curve (TRC) is an average function of n IRCs defined as

$$T(\theta) = (1/n) \sum_{j=1}^{n} P_{b_j}(\theta). \qquad [8]$$

Similarly, the group response curve (GRC) is an average function of N PRCs, that is,

$$G(b) = (1/N) \sum_{i=1}^{N} S_{\theta_i}(b). \qquad [9]$$

Illustrative PRCs and IRCs for 100 hypothetical persons were generated by randomly sampling 100 values of θ from a unit normal distribution. The resulting TRC for the simulated 100-item test is shown as the monotonically increasing function in Figure 1. The curve that is a monotonically decreasing function in Figure 1 is the PRC of $\theta = 0$, denoted by $S_0(b)$. The curve represented by ''+''s is a GRC that is obtained by taking the pointwise mean of 100 PRCs over the randomly generated 100 b values, that is, Equation 7 with N set to 100. As the number of b values approaches infinity, then $G(b)$ in Figure 1 will be a smooth monotonically decreasing curve; moreover, if the number of θ values is also very large, then $G(b)$ will be a symmetric curve of $T(\theta)$ about the vertical line of $\theta = b = 0$. For this figure, θ_i, $i = 1, 2, \ldots, 100$ and $b_j, j = 1, 2, \ldots, 100$ were randomly chosen from $N(0,1)$, so their means are not

Figure 1
Test Response Curve (...), Group Response Curve (+ + +),
and Person Response Curve (Solid Line) of the One-Parameter Logistic Model
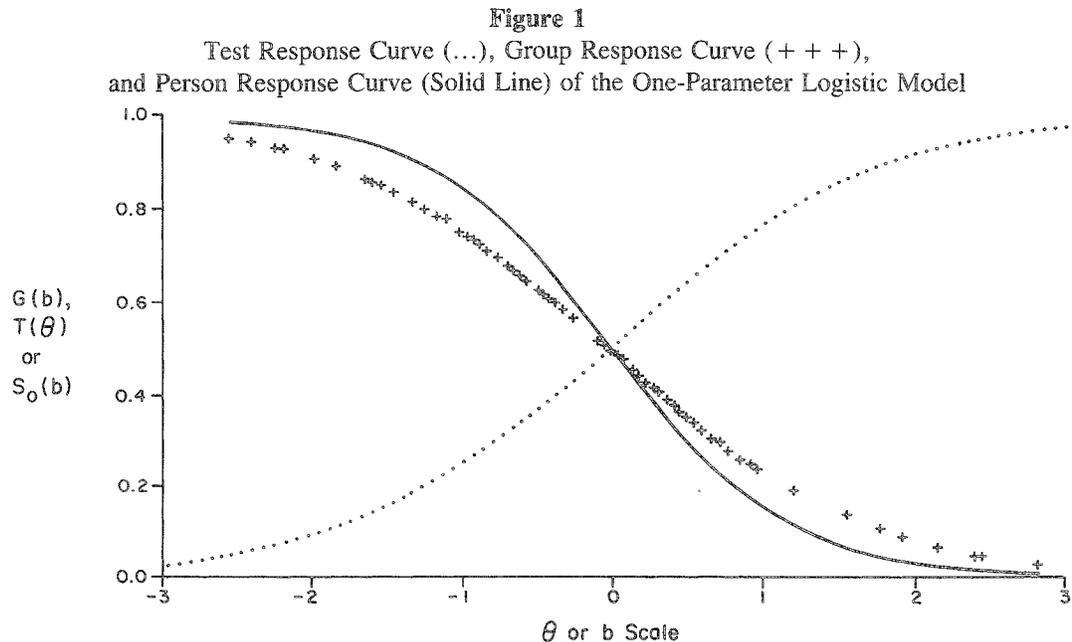
exactly zero. It can be shown numerically that $T(\theta)$ and $G(b)$ reach .5 at $\bar{\theta} = (\sum_i^{100}\theta_i)/100$ and $\bar{b} = \sum_j^{100}b_j/100$, respectively.

Let the average of $T(\theta_i)$, $i = 1, \dots, N$ be denoted by T,

$$T = (1/N) \sum_{i=1}^{N} T(\theta_i) \qquad [10]$$

and the average of $G(b_j)$, $j = 1, \dots, n$ by G,

$$G = (1/n) \sum_{i=1}^{n} G(b_j) . \qquad [11]$$

Then $T = G$, because

$$T = (1/N) \sum_{i=1}^{N} T(\theta_i)$$

$$= (1/nN) \sum_j \sum_i \{1/[1+\exp(-D(\theta_i - b_j))]\}$$

$$= (1/n) \sum_{j=1}^{n} G(b_j) = G . \qquad [12]$$

Two- and Three-Parameter Logistic Models: Problems in Person Response Curves and Group Response Curves

PRCs for the one-parameter logistic model are represented by smooth monotonically decreasing functions defined over the difficulties of the infinitely many items. For the two-parameter logistic model the PRC is no longer a smooth, monotonically decreasing curve. Figure 2 provides the graph of the PRC for the ability levels of $\theta = 0$ as well as the TRC of the two-parameter logistic model where both the ability measures, θ_1, $i = 1, 2, \dots, 100$, and the difficulties b_j, $j = 1, 2, \dots, 100$, were randomly sampled from an $N(0,1)$ distribution, while the item discrimination indices, a_j, $j = 1, \dots, 100$, were drawn from a uniform distribution within the interval (.8, 1). The TRCs and PRCs are obtained by Equation 6 with n set to 100, and

$$S_{\theta_o}(b) = \frac{1}{1+\exp[-Da(\theta_o - b)]} \qquad [13]$$

for a fixed θ_0 and variable b. The $+ + +$ line in Figure 2 is the GRC of 100 subjects. Although each PRC is locally oscillated, especially around the origin, the GRC (the mean curve of these PRCs) becomes fairly smooth and almost monotonically decreasing. Since b_j, $j = 1, \dots, 100$ are randomly selected from $N(0,1)$, a larger oscillation of PRC around the mean is expected. The GRC is expected to be smoother as the number of students and items increase to a larger number. Similar results, which are not presented here, were obtained using the three-parameter model except for the unexpected difference in the lower asymptote.

Figure 2
Test Response Curve (Solid Line), Group Response Curve (+ + +), and
Person Response Curve (Jagged Line) of the Two-Parameter Logistic Model

θ or b Scale
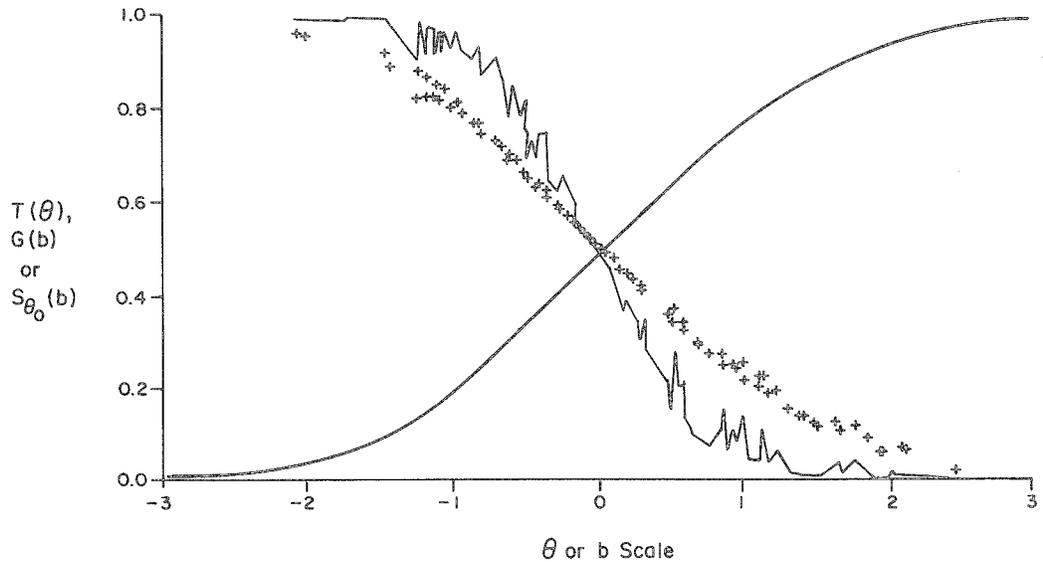
Definition of Various Extended Caution Indices

Sato's (1975) S-curve may be viewed as a discrete test response curve. The perfect S-curve divides 1's and 0's into two mutually exclusive areas with 1's under the curve and 0's above it. Note, however, that direct correspondence in this way involves a reordering of the subjects from low to high rather than from high to low, as it is typically presented by Sato and as was shown in Table 2. $T(\theta)$ represents the average probability of correctly answering items on the test when a person's ability is equal to θ. The analogy between the S-curve and a TRC may be seen by considering an alternative $N \times n$ score matrix with real numbers based on IRT rather than binary scores. More specifically, let

$$PM_{ij} = P_{b_j}(\hat{\theta}_i) \qquad [14]$$

where $\hat{\theta}_i$ is an estimated ability parameter, θ, for person i and $\hat{b}_j$ is an estimated item parameter for item j under the condition that

$$\sum_{j=1}^{n} P_{b_j}(\hat{\theta}_i) = \sum_{j=1}^{n} y_{ij} . \qquad [15]$$

Since

$$P_{b_j}(\theta_i) = S_{\theta_i}(b_j) \qquad [16]$$

for fixed i and j, the cells of the probability matrix (PM_{ij}) are also equal to $S_{\hat{\theta}_i}(\hat{b}_j)$. If the rows and columns of this matrix are arranged in the manner of the S-P Table and the columnwise sums of the cell entries are obtained, the result is N times $G(b_j)$, which corresponds to the P-curve. Similarly, n times $T(\theta_i)$ corresponding to the S-curve may be obtained by summing the cell entries for each row.

Table 4
The 127 x 32 Probability Matrix (PM_{ij}) for
Signed-Number Subtraction Problems

| Subject | Item | | | | | | $T(\hat{\theta}_i)$ | $\hat{\theta}_i$ |
	5	3	8	25	28	12		
1	.001	.001	••• .040	.002	••• .017	.082	.026	-1.242
2	.002	.012	••• .161	.031	••• .182	.341	.102	-.679
3	.002	.014	••• .174	.037	••• .207	.365	.112	-.644
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	.548	.633	••• .783	.871	••• .969	.939	.809	.410
61	.549	.635	••• .783	.872	••• .969	.939	.810	.411
62	.569	.648	••• .789	.878	••• .970	.941	.817	.424
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
116	.994	.986	••• .970	.997	••• .999	.995	.981	1.221
117	.994	.986	••• .970	.997	••• .999	.995	.981	1.221
118 *	1.000	1.000	••• 1.000	1.000	••• 1.000	1.000	1.000	+∞
$G(\hat{b}_j)$	.521	.565	.691	.702	.827	.831		
$\hat{b}_j$	.378	.289	-.066	.021	-.347	-.467		

*Nine more students answered all items correctly

Selected rows and columns of a probability matrix (PM_{ij}) are illustrated in Table 4 for a 32-item test involving the subtraction of signed numbers that was administered to a sample of 127 students (Tatsuoka & Tatsuoka, 1981). Also shown in Table 4 are the values of the estimated item and ability parameters in the two-parameter logistic model and the test and group response curves evaluated at those estimated parameter values (i.e., $T(\theta_i)$ and $G(b_j)$, respectively).

Before introducing the extended caution index, it is useful to compare the S and P curves for the data from which the estimates in Table 4 were obtained with their counterparts, i.e., *n* times $T(\theta_i)$ and *N* times $G(b_i)$. The two comparisons, S with $nT(\theta_i)$ and P with $NG(b_j)$, are provided in Figures 3 and 4, respectively. The tic marks on the curve in Figure 3 indicate the location of the θ's for the 127 students in the study. The tic marks in Figure 4 show the values of b_j for the 32 items. The close correspondence between the two pairs of curves is apparent. The number of items and the limited range of values that $\hat{b}_j$ assumes for these data obviously limits the evaluation of the correspondence between the curves in Figure 4, however.

Given the parallels between the S-P curves and the GRC and TRC, the extension of the caution index for use with the latter curves is relatively straightforward. There are, however, several natural ways in which the extension can be made. Possibly the most obvious extension is to simply replace the term ($M_{ij}^s - P_{i.}$) in the denominator of Equation 1 by its counterpart from the PM_{ij} matrix, i.e.,

$$[PM_{ij} - T(\hat{\theta}_i)] = [S_{\hat{b}_j}(\hat{\theta}_i) - T(\hat{\theta}_i)] . \qquad [17]$$

With the above substitution, the first extended caution index, $ECI1_i$, is defined:

$$ECI1_i = 1 - \frac{\sum\limits_{j}(y_{ij} - P_{i.})(y_{.j} - P_{..})}{\sum\limits_{j}[S_{\hat{\theta}_i}(\hat{b}_j) - T(\hat{\theta}_i)](y_{.j} - P_{..})} . \qquad [18]$$

Figure 3
Comparison of S-Curve (Converted to Proportion Correct by Subject)
with the Test Response Curve for the Data in Table 3
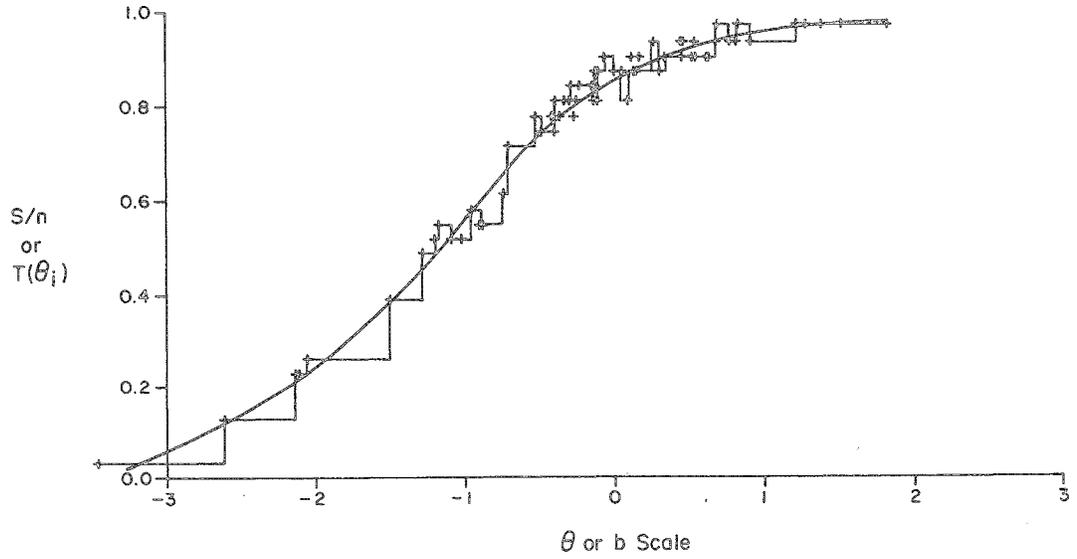
Figure 4
Comparison of P-Curve (Converted to Proportion Correct by Item)
with the Group Response Curve for the Data in Table 3
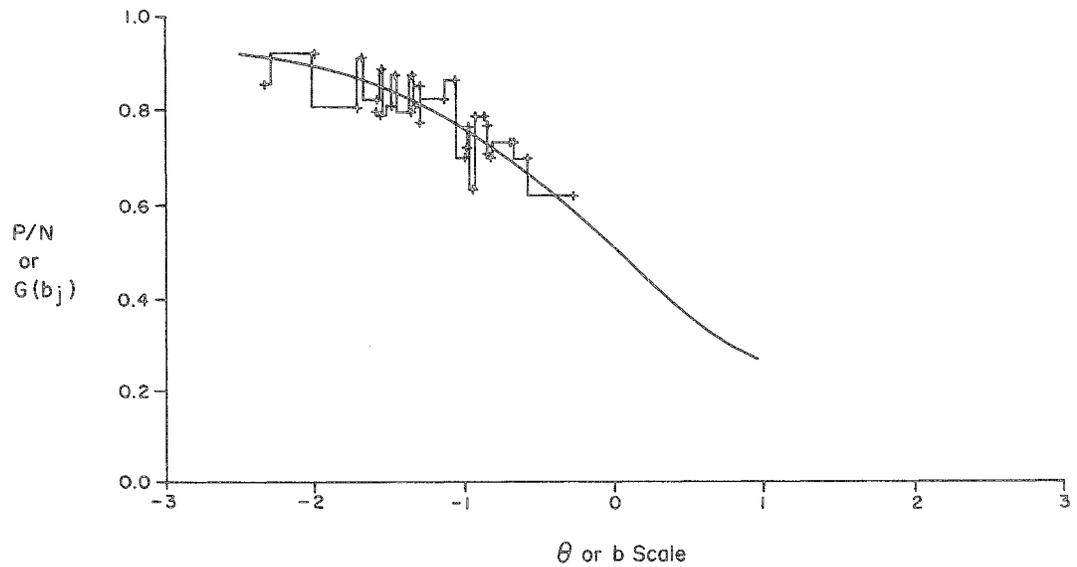
Since the maximum likelihood procedure of estimating parameters gives the relationship

$$y_{.j} = N \sum_{j=1}^{n} G(\hat{b}_j) \ . \tag{19}$$

the second extended caution index, ECI2_i, is obtained by replacing the term $(y_{.j} - P_{..})$, which appears in the numerator and the denominator of Equation 18, by $[G(\hat{b}_j) - G]$ and is defined as

$$\text{ECI2}_i = 1 - \frac{\sum_{j=1}^{n} (y_{ij} - P_{i.})[G(\hat{b}_j) - G]}{\sum_{j=1}^{n} [S_{\hat{\theta}_i}(\hat{b}_j) - T(\hat{\theta}_i)][G(\hat{b}_j) - G]} \ . \tag{20}$$

The second term is the ratio of two covariances, $y_i - P_i$ in the raw data matrix and $G - T$ from Table 4 as the numerator, and $S_{\theta_i}(b_j) - T(\theta_i)$ and $G - T$ as the denominator. After carrying out subtraction in Equation 20 the numerator becomes the covariance of the residual vector $[S_{\theta_i} - T(\theta_i)] - (y_i - P_i)$ and $G - T$.

The third natural extension of the caution index, ECI3_i, is given by the complement of the correlation ratio:

$$\text{ECI3}_i = 1 - \frac{\dfrac{\sum_{j=1}^{n} (y_{ij} - P_{i.})[G(\hat{b}_j) - G]}{\hat{\sigma}_j(y_{ij}) \, \hat{\sigma}_j[G(\hat{b}_j)]}}{\dfrac{\sum_{j=1}^{n} [S_{\hat{\theta}_i}(\hat{b}_j) - T(\hat{\theta}_i)][G(\hat{b}_j) - G]}{\hat{\sigma}_j[S_{\hat{\theta}_j}(\hat{b}_j)] \, \hat{\sigma}_j[G(\hat{b}_j)]}} \ . \tag{21}$$

Index ECI2_i takes a group-oriented vector G as a common vector in the two covariances, but G can be replaced by S_{θ_i}, a vector representing a fixed level θ_i, theoretically. This yields the numerator of the fourth index, ECI4_i, and is a measure of the relationship between an observed y_i and the theoretical vector, S_{θ_i}, at the level of θ_i. The denominator is a standardized measure, covariance of $G - T$ and $S_{\theta_i} - T(\theta_i)$,

$$\text{ECI4}_i = 1 - \frac{\sum_{j=1}^{n} (y_{ij} - P_{i.})[S_{\hat{\theta}_i}(\hat{b}_j) - T(\hat{\theta}_i)]}{\sum_{j=1}^{n} [G(\hat{b}_j) - G][S_{\hat{\theta}_i}(\hat{b}_j) - T(\hat{\theta}_i)]} \ . \tag{22}$$

Taking the correlation of both the numerator and denominator in Equation 22 gives the fifth ECI, ECI5_i:

$$\text{ECI5}_i = 1 - \frac{\dfrac{\sum_{j=1}^{n} (y_{ij} - P_{i.})[S_{\hat{\theta}_i}(\hat{b}_j) - T(\hat{\theta}_i)]}{\hat{\sigma}_j(y_{ij}) \, \hat{\sigma}_j[S_{\hat{\theta}_i}(\hat{b}_j)]}}{\dfrac{\sum_{j=1}^{n} [G(\hat{b}_j) - G][S_{\hat{\theta}_i}(\hat{b}_j) - T]}{\hat{\sigma}_j[G(\hat{b}_j)] \, \hat{\sigma}_j[S_{\hat{\theta}_i}(\hat{b}_j)]}} \ . \tag{23}$$

The denominators of the four indices are considered as normalizing constants, but the characteristics of the numerators can be divided into two categories. The indices in the first category, $ECI2_i$ and $ECI3_i$, give measures that are more group dependent because they are the covariances of the corresponding elements of the observed vector (y_{ij}) and the row-sum total vector ($y_{.j}$) and group response curve $G(b_j)$, respectively. They measure the relationship of an observed response pattern for person i to a normed variable derived from the group to which person i belongs. Thus, these indices have a similar function to the norm conformity index, NCI, defined in Tatsuoka and Tatsuoka (1982) and are conceptually similar to Donlon and Fischer's (1968) personal biserial index. The remaining two indices, $ECI4_i$ and $ECI5_i$ are more individually oriented. That means the quantities obtained from $ECI4_i$ and $ECI5_i$ reflect the extent to which a person i's response pattern (y_{ij}) relates to a theoretically derived PRC at a fixed level of θ_i. The relationship between a response pattern and the theoretically derived PRC was previously used by Trabin and Weiss (1979) for their obtained versus theoretical chi-squares. These two indices, $ECI4_i$ and $ECI5_i$, are also similar to the individual consistency index (ICI; Tatsuoka & Tatsuoka, 1982).

These extended caution indices for person i can be easily altered so that the same set of extended caution indices can be computed for item j.

Note that the arrangement of rows and columns according to the order of the proportion correct (p values) for n items and the total scores for N subjects is essential to determine S-P curves, and the values of M_{ij}^p and M_{ij}^s, $i = 1, 2, ..., N$, $j = 1, ..., n$. With the extended caution indices, the arrangement of rows and columns in monotonic order of probability is no longer necessary.

Application of New Indices for the Detection of Anomalous Responses

There is evidence that student errors on certain types of arithmetic problems are frequently quite systematic (Birenbaum & Tatsuoka, 1980; Brown & Burton, 1978; Davis & McKnight, 1980). That is, students seem to consistently apply erroneous algorithms in attempting to answer a problem of a particular form. Sometimes erroneous or incomplete rules result in the correct answer. For example, a student who consistently treats a multiplication sign as if it were an addition sign would get the correct answer to the problem $2 \times 2 = 4$, but would get it correct for the wrong reason. A score of 0 for using the wrong operation would be a better reflection of the student's ability to multiply than a score of 1 for answering "4" to the item.

Birenbaum and Tatsuoka (1980) have demonstrated that the customary 0–1 scoring of incorrect and correct answers can give the appearance of higher dimensionality and cause difficulty in attempting to apply IRT when students consistently apply erroneous rules to the addition and subtraction of signed numbers. The difficulties result from the fact that several erroneous rules frequently yield the correct answer for some problems. Correct answers for the wrong reasons not only cause problems in applying IRT, but more importantly, they can result in misleading scores and make it difficult to diagnose what the student is doing wrong.

Tatsuoka and her colleagues (Birenbaum, 1981; Birenbaum & Tatsuoka, 1980) identified several erroneous rules that were consistently applied by certain students. Birenbaum and Tatsuoka (1980) reanalyzed their data after converting 1's to 0's for items that students answered correctly for the wrong reasons. That is, an item score was changed from 1 to 0 if (1) a student was identified as consistently applying an erroneous rule and if (2) application of that erroneous rule would lead to the correct answer for the particular item in question. Analysis of the resulting modified data indicated that the data were more nearly unidimensional, and there was good evidence that IRT was more applicable to the modified data than to the original data.

Anomalous response patterns can sometimes be found by conducting an intuitive error analysis or by clinical interviews. Both approaches require enormous effort. Brown and Burton (1978) and Tatsuoka

Birenbaum, Tatsuoka, and Baillie (1980) have developed computerized approaches to error analysis. These methods, however, are expensive and are based on extensive work with highly specific item content.

Tatsuoka and Tatsuoka (in press) demonstrated an index called the individual consistency index (ICI), which was shown to be useful in detecting a variety of erroneous rules of operation with signed-number addition and subtraction problems. Using the ICI to detect examinees who are apt to have a misconception saves considerable effort because only examinees so identified have their item responses routed to the detailed error-diagnostic system. Application of the ICI is limited, however, because it requires repeated measures (i.e., several items based on an identical item form) within the test. Such repetition is not common on most tests.

The index similar to ICI, $ECI4_i$, not only avoids the repeated measure limitation but is also effective for purposes of detecting anomalous response patterns resulting from the consistent application of erroneous rules. The 13 "bugs" identified by the error-diagnostic system, SIGNBUG, were randomly chosen from the list of erroneous rules given in Tatsuoka and Tatsuoka (1981). Three of them remained in the modified data set without any changes in response patterns. The estimated item and person ability parameters needed to complete the extended caution indices were obtained from the modified data set by the computer program GETAB (Baillie, 1979).

The frequency distribution for $ECI4_i$ is displayed in Figure 5. Ten erroneous rules not existing in the modified data set are represented by black areas and the other three rules are represented by shaded areas. Three white areas on the extreme minus scale and one on the positive were examined closely again. It turned out that all of them possessed one or two misconceptions, but their application was not as consistent as that of generated bugs and changed one from another during the test.

As can be seen in Figure 5, $ECI4_i$ distinguishes persons who are consistently using erroneous rules from those who are not, provided that these erroneous rules are not popular in the data used for estimating item and person parameters. Selection of the correct data set, such as the modified data set used above, is very important when applying these new indices.

Figure 5
A Histogram of Extended Caution Index, $ECI4_i$, Obtained from
the Modified November Data (White Area) and 10 "Bugs" (Black Area),
and 3 "Left-Over Bugs" (Shaded Area)
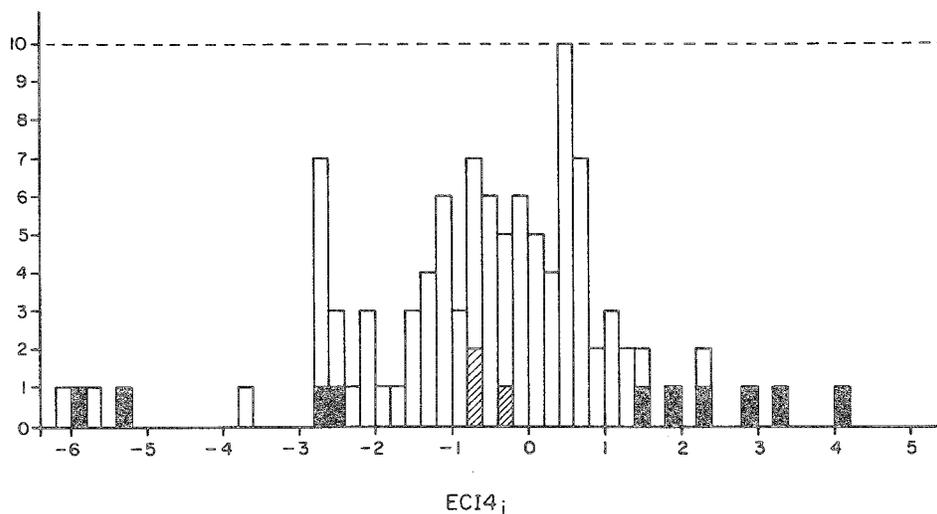
ECI4$_i$

Figure 6
A Scatterplot of Extended Caution Index, ECI4$_i$, and Total Scores ($r = .3$)
13 "Bugs" and the Modified November Data Set (Not Numbered)
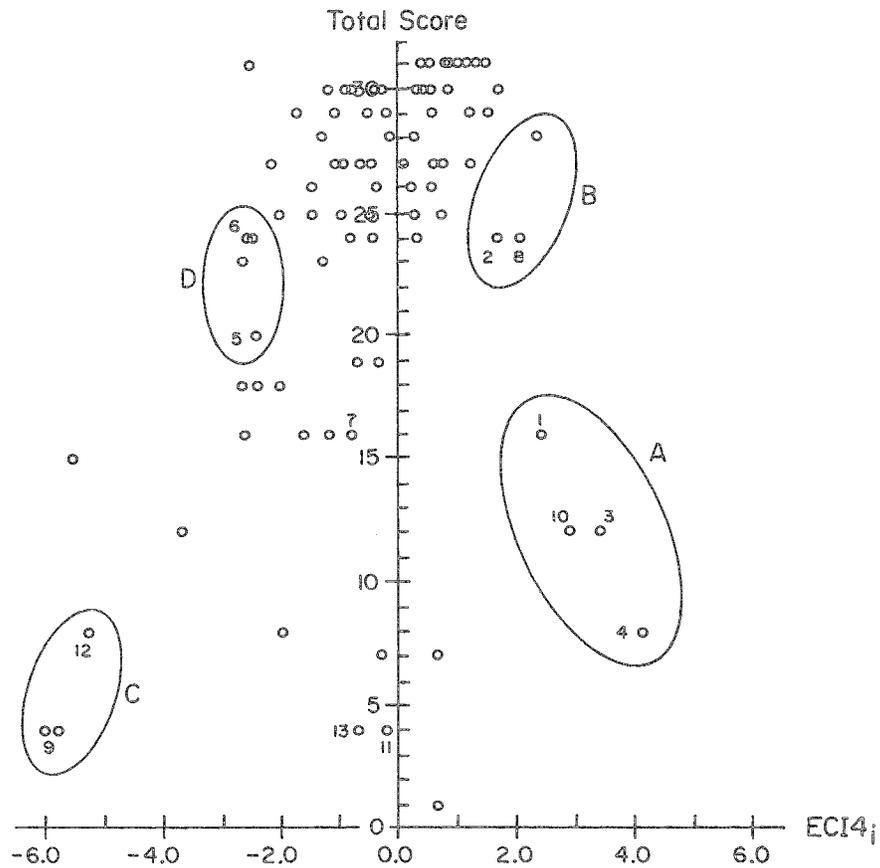
Figure 6 represents the relationship between total scores and ECI4$_i$. The bugs are numbered and roughly divided into four groups—A, B, C, and D. It is of interest to note that the members of each group have very similar response patterns. Moreover, their descriptions of errors suggest that their origin of misconceptions are initiated from the same sources. For example, the members of Group A failed to recognize the hidden plus signs of the second numbers among the test items.

As might be expected from a comparison of the algebraic expression of the coefficients ECI2$_i$ and ECI3$_i$ to ECI4$_i$, the former may be useful for other tasks for which the NCI or van der Flier's (1977) index have been found to be useful. The latter index ECI4$_i$, however, is quite effective for purposes of detecting persons who make consistent errors.

Discussion

As was shown above, the caution index which Sato (1975) developed, based solely on a comparison of observed item responses to group responses, can be readily extended to theory-based estimates of person and group response probabilities. The caution index is a linear transformation of the covariance

of a person's response pattern with one or another theoretical curves computed using IRT. Alternatively, the extended caution indices may be viewed as linear transformations of the covariance or correlation between a person's response pattern and a theoretical curve (either the PRC, as in the case of $ECI4_i$ and $ECI5_i$ or the GRC, as in the case of $ECI2_i$ or $ECI3_i$).

The application of the extended caution indices that were introduced in this paper provided strong evidence that the indices depending on the distance between a person's response pattern and his/her theoretical PRC (i.e., $ECI4_i$ and $ECI5_i$) are quite effective for purposes of identifying persons who consistently use an erroneous rule in answering signed-number arithmetic problems. This is a potentially important result that deserves further investigation with other data sets involving different types of achievement test data. If additional research yields similar results, these indices may have considerable instructional utility, because instruction can be made much more specific once it is determined that a student is consistently making an error as the result of a particular misconception.

References

Baillie, R. *GETAB: A computer program for estimating item and person parameters of the one- and two-parameter logistic model on the PLATO® system*. Urbana IL: University of Illinois, Computer-based Education Research Laboratory, 1980.

Birenbaum, M. *Error analysis—it does make a difference*. Doctoral Dissertation, University of Illinois at Urbana-Champaign, 1981.

Birenbaum, M., & Tatsuoka, K. K. *The use of information from wrong responses in measuring students' achievement* (Research Report 80-1 ONR). Urbana IL: University of Illinois, Computer-based Education Research Laboratory, 1980. (NTIS No. AD A097715)

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

Davis, R. B., & McKnight, C. The influence of semantic content on algorithmic behavior. *The Journal of Mathematical Behavior*, 1980, *3*, 39–87.

Donlon, T. F., & Fischer, F. F. An index of an individual's agreement with group determined item difficulties. *Journal of Educational and Psychological Measurement*, 1968, *28*, 105–113.

Harnisch, D. L., & Linn, R. L. Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *The Journal of Educational Measurement*, 1981, *3*, 39–87.

Levine, M. V., & Rubin, D. B. Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 1979, *4*, 269–290.

Lord, F. M. *Application of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum, 1980.

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

Lumsden, J. Tests are perfectly reliable. *British Journal of Mathematical and Statistical Psychology*, 1978, *31*, 19–26.

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

Tatsuoka, K. K., & Tatsuoka, M. M. *Detection of aberrant response patterns and their effect on dimensionality* (Research Report 80-4-ONR). Urbana IL: University of Illinois, Computer-based Education Research Laboratory, 1980.

Tatsuoka, K. K., & Tatsuoka, M. M. *Spotting erroneous rules of operation by the individual consistency index* (Research Report 81-4-ONR). Urbana IL: University of Illinois, Computer-based Education Research Laboratory, 1981.

Tatsuoka, K. K., & Tatsuok, M. M. Detection of aberrant response patterns. *Journal of Educational Statistics*, 1982, *7*, 215–231.

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 IL: University of Illinois, Computer-based Education Research Laboratory, 1980.

Tatsuoka, K. K., & Tatsuoka, M. M. Detection of aberrant response patterns. *Journal of Educational Statistics*, 1982, *7*, 215–231.

Trabin, T. E., & Weiss, D. J. *The person response curve: Fit of individuals to item characteristic curve models* (Research Report 79–7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1979.

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.

Wright, B. D. Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 1977, *14*, 97–116.

Wright, B. D., & Stone, M. H. *Best test design: Rasch measurement*. Chicago: Mesa Press, 1979.

Author's Address

Send requests for reprints or further information to Kikumi Tatsuoka, Computer-Based Educational Research Laboratory, 225 Engineering Research Laboratory, University of Illinois, 103 South Mathews, Urbana IL 61801, U.S.A.