Adaptive Estimation When the Unidimensionality Assumption of IRT is Violated

Valerie Greaud Folk
Syracuse University

Bert F. Green
Johns Hopkins University

This study examined some effects of using a unidimensional IRT model when the assumption of unidimensionality was violated. Adaptive and nonadaptive tests were formed from two-dimensional item sets. The tests were administered to simulated examinee populations with different correlations of the two underlying abilities. Scores from the adaptive tests tended to be related to one or the other ability rather than to a composite. Similar but less disparate results were obtained with IRT scoring of nonadaptive tests, whereas the conventional standardized number-correct score was equally related to both abilities. Differences in item selection from the adaptive administration and in item parameter estimation were also examined and related to differences in ability estimation. *Index terms: ability estimation, adaptive testing, item parameter estimation, item response theory, multidimensionality.*

In computerized adaptive testing (CAT), the ability (θ) of each examinee is estimated from individually tailored tests. Currently, most CAT procedures are based on unidimensional item response theory (IRT) models which incorporate a single θ to account for test performance in a particular domain. In reality, many tests measure multidimensional domains. It is not clear what the effects on θ estimation will be if such tests are administered adaptively.

The unidimensionality requirement would be less critical if multidimensional IRT models could be used in practical testing. Much work has been done to develop multidimensional IRT models (e.g., Bloxom & Vale, 1987; Bock & Aitkin, 1981; Bock & Lieberman, 1970; McKinley & Reckase, 1983a, 1983b; Reckase, 1985a, 1985b; Samejima, 1974; Sympson, 1978), but development of these models is not yet complete and practitioners are currently limited to unidimensional models.

It has been shown that unidimensional three-parameter procedures for *nonadaptive* θ estimation tend to emphasize one ability as the dataset deviates from unidimensionality. Reckase (1979) generated multidimensional datasets to fit a factor-analytic model with two or more factors and analyzed $\hat{\theta}$ values from the parameter estimation procedure (i.e., these were not adaptive estimates). For datasets with independent item subsets, $\hat{\theta}$ was most strongly related to a single factor. For datasets with a dominant first factor, $\hat{\theta}$ was most strongly related to the dominant factor.

Drasgow and Parsons (1983) conducted similar comparisons for hierarchical datasets with a single general factor and five specific factors. As correlations between the specific factors decreased, the strength

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 13, No. 4, December 1989, pp. 373–389

0146-6216/89/040373-17$2.10

of the general factor decreased. Consequently, nonadaptive $\hat{\theta}$ became more weakly correlated with the general factor and more strongly correlated with the strongest specific factor.

Ansley and Forsyth (1985) explored the characteristics of $\hat{\theta}$ from two-dimensional data produced by a noncompensatory model. Ansley and Forsyth found that as $r_{\theta_1\theta_2}$ decreased, nonadaptive $\hat{\theta}$ became more strongly correlated with θ_1 and became less correlated with θ_2 and with the average of θ_1 and θ_2. It should be noted, however, that Dimension 1 was dominant over Dimension 2 ($\bar{a}_1 = 1.23$, $\bar{a}_2 = .49$); emphasis of θ_1 in the $\hat{\theta}$ would be expected in this situation.

Way, Ansley, and Forsyth (1988) also examined $\hat{\theta}$ from two-dimensional data, explicitly comparing compensatory and noncompensatory models of data generation. For both models, they found that as $r_{\theta_1\theta_2}$ decreased, nonadaptive $\hat{\theta}$ became more strongly correlated with θ_1 and more weakly correlated with θ_2 and with the average of θ_1 and θ_2. As in the Ansley and Forsyth (1985) study, Dimension 1 was dominant over Dimension 2 in the datasets generated by both models.

The primary purpose of this study was to examine estimation of θ in adaptive tests constructed with a unidimensional model when the actual item responses were simulated from a multidimensional compensatory model. Characteristics of item selection and parameter estimation were also examined. Simulations were chosen to permit explicit comparison between estimated and true ability values and item parameters.

Method

Administrations of two types of item sets to several examinee samples were simulated using three testing methods: adaptive administration with IRT ability estimation (CAT), nonadaptive administration with IRT ability estimation (IRT), and conventional nonadaptive administration with standardized number-correct scoring of ability (CONV). The two item sets were two-dimensional either between- or within-items.

IRT Models

The unidimensional three-parameter logistic model developed by Birnbaum (1968) was chosen for the item parameter and ability estimations in the CAT and IRT administrations. This model defines the probability of a correct response to an item, $x_i = 1$, for any ability level, θ, as a function of three item parameters, a, b, and c, according to

$$P(x_i = 1|\theta) = c_i + \frac{1-c_i}{1+\exp[-1.7a(\theta-b_i)]} \quad . \tag{1}$$

A compensatory multidimensional model was chosen for this study. A compensatory model assumes that a greater degree of ability on one dimension can offset a lesser degree of ability on another dimension, whereas a noncompensatory model assumes that a certain degree of ability is required on both dimensions. The following compensatory two-dimensional model was used:

$$P(x_i = 1|\theta_1,\theta_2) = c_i + \frac{1-c_i}{1+\exp\{-1.7[a_{i1}(\theta_1-d_i)+a_{i2}(\theta_2-d_i)]\}} \quad . \tag{2}$$

Hattie (cited in Ansley & Forsyth, 1985) is credited with proposing this model, which is similar to models studied by Drasgow and Parsons (1983), McKinley and Reckase (1983b), and Reckase (1979). Equation 2 is an extension of the unidimensional model presented in Equation 1. All symbols for the two-dimensional model represent the same parameters as for the unidimensional model, except that each examinee has two ability parameters (θ_1 and θ_2) and each item has two discrimination parameters (a_1 and a_2) and a single multidimensional difficulty parameter d. In two-dimensional space, the combination of a_1 and a_2 dictates the direction of each item and d indicates the distance of each item in the defined direction.

The natural extension of the unidimensional model might seem to include two difficulties, one on each dimension. But it is easily seen that for each item, the constant in the denominator is a single value; thus the two difficulties cannot be separately identified.

The compensatory model follows from the theory of factor analysis, in which underlying characteristics account for common similarities among the variables. This model is also consistent with most of the multidimensional models currently being developed.

Structure of the Item Sets

Multidimensional between-items. A set of items will be called multidimensional between-items if it contains mutually exclusive subsets of items, such that one ability is required for one subset of items and another ability is required for another subset. For example, a general science test may contain some biology and some physics items; different abilities are needed for each subset. This structure is similar to the simulated two-dimensional item set used by Reckase (1979).

In the present study, item set BT was two-dimensional between-items. It contained 96 items, subdivided so that half of the items exclusively measured Dimension 1 and half exclusively measured Dimension 2. The 48 items measuring Dimension 1 were assigned a positive a_1 value ($\bar{a}_1 = 1.5$, $\sigma_{a_1} = .25$) and $a_2 = 0$. The 48 items measuring Dimension 2 were assigned $a_1 = 0$ and a positive a_2 value ($\bar{a}_2 = 1.5$, $\sigma_{a_2} = .25$).

The d parameters were assigned at equal intervals in the range -2.0 to $+2.0$ ($\bar{d} = 0$, $\sigma_d = 1.20$) for each subset of items.

Multidimensional within-items. A set of items will be called multidimensional within-items if every item measures some combination of several abilities and the component abilities are nonproportionately weighted, that is, if each item requires differing amounts of each ability.[1] This structure is similar to the simulated two-dimensional datasets of Ansley and Forsyth (1985) and Way et al. (1988). Hierarchical models, such as those presented by Drasgow and Parsons (1983), would also fall in this category, but a strong general factor could minimize within-item multidimensionality characteristics.

Item set WT was two-dimensional within-items. It contained 98 items divided into seven subsets, each with a different proportion of a_1 to a_2. Because the average of the two abilities is later used as a target, as recommended in Green (1988), the relationship between the dimensions was required to be additive; thus a parameters were assigned so that the mean of the sum of a_1 and a_2 ($a_T \equiv a_1 + a_2$) was approximately equal to 1.5 for each subset. Means and standard deviations of the a parameters are shown in Table 1.

Means and standard deviations of a_T are also shown. Subsets 1, 2, and 3 emphasize Dimension 1 to varying degrees, subset 4 represents both dimensions equally, and subsets 5, 6, and 7 emphasize Dimension 2 to varying degrees. The d parameters were assigned at equal intervals in the range -2.0 to $+2.0$ ($\bar{d} = 0$, $\sigma_d = 1.33$) for each subset of items.

Procedure

First, unidimensional item parameter estimates were obtained after each complete item set was

[1]Proportional weighting will not be considered here because, as Reckase, Ackerman, and Carlson (1988) have recently demonstrated, it would be undetectable in real tests and can be considered practically as a single composite of the abilities.

Table 1
Mean and Standard Deviation of True a_1, a_2, and a_T in Item Set WT, for Each Subset of Items

	Subset						
Statistic	1	2	3	4	5	6	7
a_1							
Mean	1.54	1.21	1.00	.77	.49	.25	0
SD	.26	.22	.16	.13	.09	.04	0
a_2							
Mean	0	.24	.50	.77	.97	1.25	1.54
SD	0	.04	.08	.13	.17	.20	.26
a_T							
Mean	1.54	1.46	1.50	1.54	1.46	1.50	1.54
SD	.26	.26	.24	.26	.26	.24	.26

administered to examinee samples. Then, administrations of the three testing methods to new samples of examinees were simulated, using the estimated parameters to calculate $\hat{\theta}$ in the IRT and CAT administrations. In all simulations, examinee responses were generated from the two-dimensional model but the unidimensional model was used to obtain estimates of item parameters and examinee abilities. Hence the data were multidimensional but the test administrations assumed unidimensionality.

Parameter estimation. A modified maximum likelihood procedure was used to obtain item parameter estimates. Lord's (1980) maximum likelihood equations were used to obtain a and b item parameter estimates, but Bayesian modal estimation was used to estimate $\hat{\theta}$ (Mislevy, 1986; Swaminathan & Gifford, 1986). The Bayesian modal formula is but a slight addition to the maximum likelihood formulas; the function being maximized is the log likelihood plus the prior distribution density. The method may be considered a compromise between the LOGIST procedures (Wingersky, 1984) and the Bayesian methods (Bock & Aitkin, 1981). All methods give satisfactory results. Because the Bayesian modal estimator is used in the CAT procedure in scoring examinees, it seemed wise to include it in the item parameter estimation process.

The estimation process required initial $\hat{a}$, $\hat{b}$, and $\hat{\theta}$ values as input. Equations from Hambleton and Swaminathan (1985, pp. 144–147) were used to calculate starter values for each item's $\hat{b}$ and each examinee's $\hat{\theta}$. Initial $\hat{a}$ values were set at 1.0. The parameter estimation procedure recalculated, in turn, $\hat{a}$ and $\hat{b}$, then $\hat{\theta}$, until the average squared change in estimation was less than .0001 for $\hat{a}$, $\hat{b}$, and $\hat{\theta}$.

Examinee samples for parameter estimation. Thirteen samples of simulated examinees (simulees) were generated. Each sample contained 2,000 simulees, with two θ values selected for each simulee from a bivariate distribution ($\bar{\theta} = 0$, $\sigma_\theta = 1.0$) with one of five possible degrees of correlation between θ_1 and θ_2: $r_{\theta_1\theta_2} = .20, .40, .60, .80$, or 1.00. An original complete set of five samples was drawn where $r_{\theta_1\theta_2}$ for each sample was a different value of the five possible correlations. Then, in order to examine the effects of sampling, four partial replicate sets of samples were drawn from populations with $r_{\theta_1\theta_2} = .20$ and $r_{\theta_1\theta_2} = .60$.

For the simulated tests, 13 additional simulee samples were drawn corresponding to the original and replication samples described above for the parameter estimation.

CAT administration. Item sets BT and WT were each administered adaptively to the simulee samples and unidimensional $\hat{\theta}$ values were obtained. Each simulee began with $\hat{\theta} = 0$ and then received 15 items which were successively selected to have the highest information value relative to the current (unidimensional) $\hat{\theta}$ of that simulee. A provisional $\hat{\theta}$ was calculated after each item response using Owen's

(1975) approximation method. The final $\hat{\theta}$, after 15 items, was obtained using a Bayesian modal method. The final $\hat{\theta}$ values were recorded for each simulee as the unidimensional model's best $\hat{\theta}$; they are referred to as $\hat{\theta}_{CAT}$.

IRT administration. Complete item sets BT and WT were also administered nonadaptively. IRT-based $\hat{\theta}$ values were then calculated; these are referred to as $\hat{\theta}_{IRT}$.

Conventional administration. In the conventional tests, complete item sets BT and WT were again administered, but $\hat{\theta}$ values were estimated by calculating standardized number-correct scores. These are referred to as $\hat{\theta}_{CONV}$.

Analyses

For all administration methods, the relationships between $\hat{\theta}$ and each of the true abilities, θ_1 and θ_2, were examined by calculating the multiple correlation of $\hat{\theta}$ with θ_1 and θ_2 ($R_{1,2}$). Bivariate correlations between $\hat{\theta}$ and θ_1 (r_1) and between $\hat{\theta}$ and θ_2 (r_2) were also calculated and the absolute difference between r_1 and r_2 ($|r_1 - r_2|$) was calculated. In addition, because Green (1988) recommended examination of the relationship between $\hat{\theta}$ and θ_{avg} $[(\theta_1 + \theta_2)/2]$, the bivariate correlation between $\hat{\theta}$ and θ_{avg} (r_{avg}) was calculated.

Adaptive item selection. The items selected by CAT were examined to assess the representation of the two dimensions. When a CAT was administered to a simulee sample, 15 items were selected for each of 2,000 simulees, for a total of 30,000 item selections. The number of occasions on which items from each subset were selected was divided by 30,000 to indicate the proportion of occasions on which items from the subset were selected for administration. Item sets BT and WT differed in the number of subsets; thus the item selection analyses were conducted slightly differently for each.

For item set BT, the subset administration proportions p_1 and p_2 indicated the representation of each dimension because the two dimensions were mutually exclusive. The absolute difference $|p_1 - p_2|$ was also calculated as an index of overemphasis of a dimension.

Item set WT contained seven subsets of items, each with different relative weights of Dimension 1 and Dimension 2; administration proportions p_{S1}, p_{S2}, ..., p_{S7} were calculated for each of the subsets. However, because each subset contained different weightings of the two dimensions, the subset administration proportions were not useful for determining whether the dimensions themselves were proportionately represented in the CATs. To assess the representation of each dimension, two groupings of the items were created: The Dimension 1 group contained items from subsets which measured Dimension 1 more than Dimension 2 (subsets 1, 2, and 3); the Dimension 2 group contained items from subsets which measured Dimension 2 more than Dimension 1 (subsets 5, 6, and 7). Subset 4 was excluded from this comparison because these items measured both dimensions equally. Administration proportions p_{D1} and p_{D2} were then totaled for the two dimension groups.

Difficulty parameter estimation. Average $\hat{b}$ and $\sigma_{\hat{b}}$ values were calculated for each set of estimations. Differences between the true and estimated difficulty parameters were assessed by calculating average absolute deviations between d and $\hat{b}$ (AAD$_{\hat{b}}$). Correlations between $\hat{b}$ and d ($r_{\hat{b}d}$) were also calculated.

Discrimination parameter estimation. The $\hat{a}$ values could not be compared directly to a single true value because there were two true parameters (a_1 and a_2) for each item. Different analyses were chosen for item sets BT and WT because they had differing structures.

For item set BT, the magnitude of a_1 for Dimension 1 items (these items had $a_2 = 0$) was the same as the magnitude of a_2 for Dimension 2 items ($a_1 = 0$). Thus the relative emphasis of each dimension was assessed by calculating the average $\hat{a}$ separately for each subset of items: the average $\hat{a}$ for the items measuring Dimension 1 [avg $\hat{a}(1)$] and the average $\hat{a}$ for items measuring Dimension 2 [avg $\hat{a}(2)$]. AADs between $\hat{a}$ and a_1 for Dimension 1 items [AAD(1)] and between $\hat{a}$ and a_2 for Dimension 2 items [AAD(2)] and the absolute difference between AAD(1) and AAD(2) [|AAD(1) − AAD(2)|] were also calculated.

In item set WT, the average $\hat{a}$ for each subset was calculated [avg $\hat{a}$(S1), avg $\hat{a}$(S2), ..., avg $\hat{a}$(S7)]; however, these values combine the true a parameters for Dimension 1 and Dimension 2 and cannot be directly compared as in item set BT. Therefore, the sum of the true values ($a_T \equiv a_1 + a_2$) was used as a standard (as recommended in Green, 1988) and AADs between $\hat{a}$ and a_T were calculated for each subset [AAD(S1), AAD(S2), ..., AAD(S7)].

The relative size of $\hat{a}$ was also assessed for the dimension groups (Dimension 1 group subsets 1, 2, and 3; Dimension 2 group subsets 5, 6, and 7). Average $\hat{a}$ values [avg $\hat{a}$(D1), avg $\hat{a}$(D2)] and AADs between $\hat{a}$ and a_T [AAD(D1), AAD(D2)] were calculated for each dimension group; the absolute difference [|AAD(D1) − AAD(D2)|] was also calculated.

Results and Discussion

Item Set BT

Ability estimation. Tables 2 and 3 summarize the results of the analyses of the ability estimates for the original and replication samples of simulees, respectively. Across simulee samples with different $r_{\theta_1\theta_2}$, comparison of $\hat{\theta}_{CONV}$ values shows no differences between r_1 and r_2. However, a striking pattern emerged for the $\hat{\theta}_{IRT}$ and $\hat{\theta}_{CAT}$ values: As $r_{\theta_1\theta_2}$ decreased, greater differences between r_1 and r_2 were consistently found, indicating that $\hat{\theta}$ was closer to either θ_1 or θ_2. This pattern was especially evident for $\hat{\theta}_{CAT}$.

In some cases, Dimension 1 was emphasized; in others, Dimension 2 was emphasized. It is suspected that this occurred because the item set was constructed to be perfectly balanced and random differences in responding upset that balance (empirical evidence of this balance is shown in the $\hat{\theta}_{CONV}$ values, which do not emphasize either dimension for any sample). If the item set had not been perfectly balanced in number and quality of items, the dimension with the better or more numerous items would have predominated—as was demonstrated in the studies of Ansley and Forsyth (1985) and Way et al. (1988).

$|r_1 - r_2|$ was included to indicate the tendency of $\hat{\theta}$ to emphasize one dimension, regardless of which dimension happened to be emphasized for a particular sample. As $r_{\theta_1\theta_2}$ decreased, $|r_1 - r_2|$ increased somewhat for $\hat{\theta}_{IRT}$ and increased dramatically for $\hat{\theta}_{CAT}$, indicating increasingly stronger relationships between $\hat{\theta}$ and either θ_1 or θ_2.

Examination of $R_{1,2}$ in Tables 2 and 3 indicates that $\hat{\theta}$ is strongly related to some optimal combination of θ_1 and θ_2—even for simulee samples with low $r_{\theta_1\theta_2}$. However, θ_1 and θ_2 are not usually known and their combination could not be discovered in real tests, hence $\hat{\theta}$ could not be meaningfully interpreted. Even if the combination could be discovered, the implications are quite different if both dimensions are equally weighted or if one dimension is emphasized.

A CAT should provide a θ estimate that is in accordance with the relative importance of the corresponding dimensions in the item pool; if the pool represents two dimensions equally, the $\hat{\theta}_{CAT}$ values should be equally related to θ_1 and θ_2. In these simulations, the dimensions were identical in the number and quality of items. If CAT had maintained the original representation, $\hat{\theta}_{CAT}$ values should have been equally related to θ_1 and θ_2—as were $\hat{\theta}_{CONV}$ values.

Item selection. Subset administration proportions are shown for the original and replication samples in Table 4. If proportional representation were achieved, 50% of the items should have been selected from each dimension. For simulee samples with high $r_{\theta_1\theta_2}$, p_1 and p_2 were very close to 50%. As $r_{\theta_1\theta_2}$ decreased, an increasingly greater proportion of items was selected from one dimension or the other. For

Table 2
Multiple Correlations Between $\hat{\theta}$ and θ_1 and θ_2 ($R_{1,2}$), Bivariate Correlations Between $\hat{\theta}$ and θ_{avg} (r_{avg}), $\hat{\theta}$ and θ_1 (r_1), and $\hat{\theta}$ and θ_2 (r_2), and $|r_1-r_2|$ from CAT, IRT, and CONV Administrations of Item Set BT to Original Samples, for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Test and Statistic	.20	.40	.60	.80	1.00
CAT					
$R_{1,2}$	.94	.92	.93	.95	.96
r_{avg}	.77	.90	.93	.95	.96
r_1	.94	.63	.80	.91	.96
r_2	.26	.87	.87	.89	.96
$\|r_1-r_2\|$	.68	.24	.07	.01	.00
IRT					
$R_{1,2}$	.96	.97	.97	.98	.98
r_{avg}	.95	.97	.97	.98	.98
r_1	.80	.78	.87	.93	.98
r_2	.67	.83	.87	.93	.98
$\|r_1-r_2\|$	.12	.04	.00	.00	.00
CONV					
$R_{1,2}$	.96	.97	.97	.97	.98
r_{avg}	.96	.97	.97	.97	.98
r_1	.75	.80	.86	.92	.98
r_2	.74	.81	.87	.92	.98
$\|r_1-r_2\|$	.01	.01	.01	.00	.00

example, the original simulee sample with $r_{\theta_1\theta_2} = .20$ received items almost exclusively from Dimension 1. In general, $|p_1 - p_2|$ increased as $r_{\theta_1\theta_2}$ decreased.

The relative sizes of p_1 and p_2 were reflected in the $\hat{\theta}_{CAT}$ values. When $r_{\theta_1\theta_2}$ was low, the CATs did not proportionately select items from the dimensions and the $\hat{\theta}_{CAT}$ values did not equally represent θ_1 and θ_2. It is not surprising that simulees in the original sample with $r_{\theta_1\theta_2} = .20$, for example, had $\hat{\theta}_{CAT}$ values that were more strongly related to θ_1 because they received a test composed primarily of items from Dimension 1.

Relation of difficulty parameters to their estimates. Analyses of the difficulty parameters for the original and replication samples are shown in Table 5. Although the $AAD_{\hat{b}}$ values indicate a slight positive bias in $\hat{b}$ as $r_{\theta_1\theta_2}$ decreased, the differences are minimal. Correlations between $\hat{b}$ and d were all above .99, indicating a strong linear relationship. These results demonstrate that the unidimensional model provided accurate $\hat{b}$ values for all simulee samples.

Note, however, that the task of the estimation procedure was simply to find a unidimensional ordering of the $\hat{b}$ values, where the true d parameters represented difficulty along a single axis in multidimensional space defined by the combination of the a parameters. This made the task somewhat easier. Also, the

Table 3
Multiple Correlations Between $\hat{\theta}$ and θ_1 and θ_2 ($R_{1,2}$), Bivariate Correlations Between $\hat{\theta}$ and θ_{avg} (r_{avg}), $\hat{\theta}$ and θ_1 (r_1), and $\hat{\theta}$ and θ_2 (r_2), and $|r_1-r_2|$ from CAT, IRT, and CONV Administrations of Item Set BT to Original (1) and Replication (2-5) Samples, for Correlation Between θ_1 and θ_2 of .20 and .60

	$r(\theta_1,\theta_2) = .20$					$r(\theta_1,\theta_2) = .60$				
Test and Statistic	1	2	3	4	5	1	2	3	4	5
CAT										
$R_{1,2}$	.94	.94	.94	.95	.94	.93	.93	.93	.94	.94
r_{avg}	.77	.76	.78	.75	.79	.93	.93	.93	.92	.92
r_1	.94	.94	.28	.21	.31	.80	.84	.84	.76	.75
r_2	.26	.25	.94	.95	.93	.87	.82	.83	.90	.90
$\|r_1-r_2\|$	.68	.69	.65	.74	.61	.07	.02	.01	.14	.15
IRT										
$R_{1,2}$	.96	.96	.95	.96	.96	.97	.98	.98	.97	.98
r_{avg}	.95	.95	.95	.95	.95	.97	.98	.98	.97	.98
r_1	.80	.81	.67	.62	.69	.87	.87	.87	.86	.87
r_2	.67	.66	.80	.84	.79	.87	.87	.88	.88	.89
$\|r_1-r_2\|$	.12	.15	.13	.22	.10	.00	.01	.00	.03	.02
CONV										
$R_{1,2}$	.96	.96	.96	.96	.96	.97	.97	.97	.97	.97
r_{avg}	.96	.96	.96	.96	.96	.97	.97	.97	.97	.97
r_1	.75	.73	.75	.74	.76	.86	.86	.87	.87	.87
r_2	.74	.75	.75	.74	.74	.87	.87	.87	.87	.87
$\|r_1-r_2\|$	.01	.02	.00	.00	.02	.01	.01	.00	.00	.00

scale of $\hat{b}$ was tied to the $\hat{\theta}$ scale. $\hat{\theta}$ values were standardized, as were the original true θ_1 and θ_2 values; thus the scale of $\hat{b}$ was the same as that of the true d values.

Relation of discrimination parameters to their estimates. Analyses of the discrimination parameters are shown in Table 6 for the original and replication samples. Average $\hat{a}$ values were highest for the sample with $r_{\theta_1\theta_2} = 1$; these values were slightly higher than the true averages, indicating a slight positive bias in the $\hat{a}$ values.

For simulee samples with higher $r_{\theta_1\theta_2}$, the AADs were virtually the same in the two subsets of items. As $r_{\theta_1\theta_2}$ decreased, greater differences were obtained between the AADs for each subset. For example, for the original simulee sample with $r_{\theta_1\theta_2} = .20$, $\hat{a}$ values were considerably lower for the items measuring Dimension 2 than for the items measuring Dimension 1.

The a parameters have two important implications. One implication is that the response of an examinee to an item with a larger $\hat{a}$ will carry greater weight in the calculation of that examinee's $\hat{\theta}$ than will that examinee's response to an item with a smaller $\hat{a}$. Indeed, this is precisely what happened when $\hat{\theta}_{IRT}$ values were calculated: The dimension receiving larger $\hat{a}$ was more strongly emphasized in simulees' $\hat{\theta}_{IRT}$ values. For example, when item set BT was administered to the original sample with $r_{\theta_1\theta_2} = .20$, the Dimension 1 items received higher $\hat{a}$ [avg $\hat{a}(1) = 1.02$, compared with avg $\hat{a}(2) = .82$ for Dimension 2 items, from

Table 4
Item Selection Proportions From CAT Administrations of Item Set BT to Original and Replication Samples for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Sample and Proportion	.20	.40	.60	.80	1.00
Original Sample					
p_1	.88	.35	.44	.53	.52
p_2	.12	.65	.56	.47	.48
$\lvert p_1-p_2\rvert$	.76	.30	.13	.06	.04
Replication 1					
p_1	.88		.50		
p_2	.12		.50		
$\lvert p_1-p_2\rvert$	.76		.00		
Replication 2					
p_1	.11		.50		
p_2	.89		.50		
$\lvert p_1-p_2\rvert$	.77		.00		
Replication 3					
p_1	.01		.36		
p_2	.99		.64		
$\lvert p_1-p_2\rvert$	.97		.28		
Replication 4					
p_1	.16		.37		
p_2	.84		.63		
$\lvert p_1-p_2\rvert$	.68		.27		

Table 6]. For these simulees, $\hat{\theta}_{IRT}$ was more strongly related to θ_1 than to θ_2 ($r_{\hat{\theta}\theta_1}$ = .80, compared with $r_{\hat{\theta}\theta_2}$ = .67, from Table 2).

A second implication is that an item with a larger $\hat{a}$ is more likely to be selected in a CAT than is an item with a smaller $\hat{a}$. This is exactly what was observed when simulee samples had weakly correlated θs. For example, when item set BT was administered to the original simulee sample with $r_{\theta_1\theta_2}$ = .20, the Dimension 1 items received larger $\hat{a}$ values and were more likely to be selected in an adaptive test (88%, compared with 12% for Dimension 2 items, from Table 4). The $\hat{\theta}_{CAT}$ values of these simulees were much more strongly related to θ_1 than to θ_2 ($r_{\hat{\theta}\theta_1}$ = .94, compared with $r_{\hat{\theta}\theta_2}$ = .26, from Table 2). In CAT, the larger $\hat{a}$ values resulted not only in an overemphasis of certain items in the calculation of θ, but also in the overselection of those items in the construction of the CAT.

Item Set WT

Ability estimation. Tables 7 and 8 contain analyses of the θ estimates for the original and replication samples of simulees. These results indicate that all methods of administering item set WT produced $\hat{\theta}$ equally related to θ_1 and θ_2.

Item selection. Subset administration proportions for the original and replication samples are shown

Table 5
Average $\hat{b}$, Standard Deviation of $\hat{b}$, and Average Absolute Deviations Between $\hat{b}$ and d (AAD) for Item Set BT, From Original and Replication Samples, for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Group and Statistic	.20	.40	.60	.80	1.00
Original Sample					
Average	.12	.07	.03	.01	−.01
SD	1.49	1.36	1.30	1.22	1.17
AAD	.26	.16	.10	.05	.04
Replication 1					
Average	.12		.00		
AAD	.27		.13		
Replication 2					
Average	.16		.06		
AAD	.28		.11		
Replication 3					
Average	.11		.04		
AAD	.31		.14		
Replication 4					
Average	.08		.02		
AAD	.26		.11		

in Tables 9 and 10. Proportional representation of the subsets was obtained if p_n equaled 14%. For most samples, items from subset 4 (which measured both dimensions equally) were administered most often.

Dimension group administration proportions are also shown in Tables 9 and 10. The dimensions were fairly evenly represented in the CATs. This is in keeping with the analyses indicating that $\hat{\theta}_{CAT}$ values for item set WT were evenly composed of θ_1 and θ_2.

Relation of difficulty parameters to their estimates. Analyses of the difficulty parameters are shown in Table 11. These results indicate that the unidimensional model provided reasonably accurate $\hat{b}$ values.

Relation of discrimination parameters to their estimates. Table 12 contains analyses of the discrimination parameter estimates obtained from the original samples. As $r_{\theta_1\theta_2}$ decreased, greater differences were found among the avg $\hat{a}$ values for the item subsets. The larger AADs tended to occur for subsets in which the weighting of one dimension was greater than the weighting of the other dimension, for example, for subsets 1, 2, 6, and 7. AADs were the smallest for subset 4, which equally weighted both dimensions, indicating that the $\hat{a}$ values for these items were closest to a_T. Items from subset 4 also tended to be selected more frequently in the CATs, as was shown in Tables 10 and 11.

Dimension group comparisons are also shown in Tables 12 and 13 for the original and replication samples. Comparison of these AADs shows no imbalance in the magnitude of $\hat{a}$ between the dimension groups. Indeed, no imbalance in the selection of items from the dimension groups was observed in the CATs.

Summary and Conclusions

This study has demonstrated that under certain conditions, use of a unidimensional model will bias

Table 6
Average $\hat{a}$ and Average Absolute Deviations Between $\hat{a}$ and a_1 (Subset 1) and Between $\hat{a}$ and a_2 (Subset 2) for Item Set BT, From Original and Replication Samples, for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Group and Statistic	.20	.40	.60	.80	1.00
Original Sample					
Average $\hat{a}$(1)	1.02	1.00	1.15	1.36	1.65
Average $\hat{a}$(2)	.82	1.12	1.24	1.42	1.64
AAD(1)	.48	.50	.35	.17	.18
AAD(2)	.68	.39	.28	.15	.17
|AAD(1)−AAD(2)|	.20	.11	.08	.02	.01
Replication 1					
Average $\hat{a}$(1)	1.04		1.13		
Average $\hat{a}$(2)	.82		1.20		
AAD(1)	.46		.37		
AAD(2)	.69		.31		
|AAD(1)−AAD(2)|	.23		.06		
Replication 2					
Average $\hat{a}$(1)	.80		1.16		
Average $\hat{a}$(2)	1.06		1.18		
AAD(1)	.70		.35		
AAD(2)	.45		.32		
|AAD(1)−AAD(2)|	.25		.04		
Replication 3					
Average $\hat{a}$(1)	.69		1.11		
Average $\hat{a}$(2)	1.16		1.22		
AAD(1)	.81		.42		
AAD(2)	.35		.29		
|AAD(1)−AAD(2)|	.46		.13		
Replication 4					
Average $\hat{a}$(1)	.81		1.12		
Average $\hat{a}$(2)	1.04		1.24		
AAD(1)	.69		.38		
AAD(2)	.48		.27		
|AAD(1)−AAD(2)|	.22		.11		

parameter estimation, adaptive item selection, and ability estimation from CAT and IRT administrations. The parameter estimation procedure appeared to lay the groundwork of this bias in the $\hat{a}$ values for item set BT. For samples with high $r_{\theta_1\theta_2}$, $\hat{a}$ values were equally related to a_1 and a_2. However, as $r_{\theta_1\theta_2}$ decreased, $\hat{a}$ increasingly emphasized either a_1 or a_2. This discrepancy was reflected in the $\hat{\theta}_{IRT}$ values: The dimensions receiving larger $\hat{a}$ exerted a stronger influence on nonadaptive $\hat{\theta}_{IRT}$ values.

The $\hat{\theta}_{CAT}$ values were led one step further astray. The $\hat{a}$ values were used not only in the estimation of $\hat{\theta}$, but also in the selection of items. More discriminating items tended to be selected and, as these tended to come from one dimension, the adaptive test as a whole became primarily a test measuring that dimension. Simulees' true θ values corresponding to the overselected dimension were necessarily the most important influence on their $\hat{\theta}_{CAT}$ values.

Table 7
Multiple Correlations Between $\hat{\theta}$ and θ_1 and θ_2 ($R_{1,2}$), Bivariate Correlations Between θ and θ_{avg} (r_{avg}), θ and θ_1 (r_1), θ and θ_2 (r_2), and $|r_1-r_2|$ from CAT, IRT, and CONV Administrations of Item Set WT to Original Samples, for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Test and Statistic	.20	.40	.60	.80	1.00
CAT					
$R_{1,2}$	.93	.94	.95	.95	.96
r_{avg}	.93	.94	.95	.95	.96
r_1	.71	.80	.84	.91	.96
r_2	.72	.77	.85	.90	.96
$\|r_1-r_2\|$	.01	.03	.01	.00	.00
IRT					
$R_{1,2}$	.97	.98	.98	.98	.98
r_{avg}	.97	.98	.98	.98	.98
r_1	.75	.82	.87	.93	.98
r_2	.75	.81	.88	.93	.98
$\|r_1-r_2\|$	.01	.01	.00	.00	.00
CONV					
$R_{1,2}$	.96	.97	.97	.97	.97
r_{avg}	.96	.97	.97	.97	.97
r_1	.74	.81	.87	.92	.97
r_2	.75	.81	.87	.93	.97
$\|r_1-r_2\|$	.01	.00	.00	.01	.00

It had initially been expected that item set WT would pose serious problems if estimation biases occurred and it became necessary to find a means of separating the dimensions, but the unidimensional model seemed to handle item set WT fairly well. Why did this item set fail to exhibit a strong emphasis of one true θ for $\hat{\theta}_{IRT}$ and $\hat{\theta}_{CAT}$? One possible explanation is that because item set WT consisted of several different combinations of the two dimensions, there was no explicit trade-off between the dimensions during parameter estimation. The parameter estimation procedure did have a tendency to emphasize the fourth subset and to underrepresent subsets in which one dimension predominated (subsets 1 and 7), but both of the dimensions retained equal influence. The $\hat{\theta}$ values maintained the balance of the two dimensions established in the parameter estimates and were equally related to θ_1 and θ_2 in all administration methods.

These results indicate the potentially adverse effects of using a unidimensional model to estimate θ from adaptive tests. Although the results are consistent with those of other researchers, this study has highlighted the effects of adaptive testing when item parameter estimates are used not only to estimate θ but also to select the items on which θ will be estimated.

It is of particular concern that CAT is strongly affected by the tendency of the discrimination parameters to emphasize one dimension. CAT may not maintain the representation of dimensions that exists in the

Table 8
Multiple Correlations Between $\hat{\theta}$ and θ_1 and θ_2 ($R_{1,2}$),
Bivariate Correlations Between $\hat{\theta}$ and θ_{avg} (r_{avg}), $\hat{\theta}$ and θ_1 (r_1),
and $\hat{\theta}$ and θ_2 (r_2), and $|r_1-r_2|$ from CAT, IRT, and CONV
Administrations of Item Set WT to Original (1) and Replication
(2-5) Samples, for Correlation of θ_1 and θ_2 of .20 and .60

	$r(\theta_1,\theta_2) = .20$					$r(\theta_1,\theta_2) = .60$				
Test and Statistic	1	2	3	4	5	1	2	3	4	5
CAT										
$R_{1,2}$	.93	.93	.93	.93	.93	.95	.94	.94	.95	.95
r_{avg}	.93	.93	.93	.93	.93	.95	.94	.94	.95	.95
r_1	.71	.68	.71	.68	.72	.84	.85	.85	.82	.83
r_2	.72	.76	.73	.76	.73	.85	.83	.85	.87	.87
$\|r_1-r_2\|$	.01	.07	.03	.08	.01	.01	.02	.00	.05	.04
IRT										
$R_{1,2}$	.97	.97	.97	.97	.97	.98	.98	.98	.98	.98
r_{avg}	.97	.97	.97	.97	.97	.98	.98	.98	.98	.98
r_1	.75	.74	.75	.74	.77	.87	.87	.88	.87	.87
r_2	.75	.76	.75	.77	.74	.88	.87	.88	.88	.88
$\|r_1-r_2\|$	.01	.01	.00	.03	.02	.00	.01	.00	.01	.01
CONV										
$R_{1,2}$	.96	.96	.96	.96	.96	.97	.97	.97	.97	.97
r_{avg}	.96	.96	.96	.96	.96	.97	.97	.97	.97	.97
r_1	.74	.75	.75	.75	.76	.87	.86	.87	.87	.88
r_2	.75	.45	.74	.74	.74	.87	.87	.87	.87	.86
$\|r_1-r_2\|$	.01	.01	.01	.00	.02	.00	.01	.00	.00	.01

complete pool of items; this shifting in the emphasis of the dimensions occurs without the awareness of the test constructor.

Some tempering conclusions should be kept in mind. For item set BT, as long as the dataset did not deviate greatly from unidimensionality, the unidimensional model provides a reasonable approximation. If $r_{\theta_1\theta_2}$ is low, the datasets are better characterized as multidimensional and they cannot be approximated with a unidimensional model, but it is likely that the dimensions would be more readily identifiable as distinct domains. Then they could be separated into different tests, or a selection rule could be incorporated to ensure equal representation of both dimensions (e.g., Thomas & Green, 1989).

These results argue strongly for the necessity of two current research trends: (1) dimensionality assessment methods to reduce reliance on content identifiability of item subsets, and (2) development of multidimensional models. (Hambleton & Rovinelli, 1986, and McDonald, 1965, 1981, 1982, have reported nonlinear factor analysis as a promising approach to developing dimensionality assessment methods.)

Item sets and examinee samples with different characteristics than those used in this study should also be examined. In particular, noncompensatory multidimensional within-item sets are perhaps more

Table 9
Item Selection Proportions for Item Subsets and Dimension Groups From CAT Administrations of Item Set WT to Original Samples, for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Item Group	.20	.40	.60	.80	1.00
Subset					
S1	.06	.11	.07	.13	.16
S2	.10	.12	.10	.13	.12
S3	.19	.19	.17	.16	.15
S4	.27	.24	.22	.22	.16
S5	.21	.18	.20	.14	.11
S6	.14	.12	.15	.12	.12
S7	.03	.05	.08	.10	.17
Dimension					
D1	.34	.41	.34	.41	.43
D2	.39	.35	.43	.36	.41

realistic than the compensatory multidimensional within-item set used in these studies. Unequal representation of dimensions should be examined because it is not likely that dimensions would be perfectly balanced among the items in a test. Item sets with more than two dimensions should also be examined. However, it is expected that the conditions simulated in this study pose the greatest challenge for unidimensional IRT models: With equal representation of two dimensions, one dimension is effectively pitted against the other as the unidimensional model attempts to find a single ordering of items and examinees.

Table 10
Item Selection Proportions for Item Subsets and Dimension Groups From CAT Administrations of Item Set WT to Original (1) and Replication (2-5) Samples, for Correlation Between θ_1 and θ_2 of .20 and .60

	$r(\theta_1,\theta_2) = .20$					$r(\theta_1,\theta_2) = .60$				
Item Group	1	2	3	4	5	1	2	3	4	5
Subset										
S1	.06	.03	.05	.04	.03	.07	.10	.07	.08	.07
S2	.10	.10	.10	.09	.13	.10	.14	.10	.09	.09
S3	.19	.21	.17	.19	.23	.17	.19	.21	.16	.15
S4	.27	.25	.24	.25	.25	.22	.22	.24	.19	.24
S5	.21	.20	.21	.19	.18	.20	.15	.17	.17	.17
S6	.14	.14	.13	.17	.12	.15	.10	.14	.14	.15
S7	.03	.06	.09	.08	.06	.08	.09	.08	.16	.14
Dimension										
D1	.34	.34	.32	.31	.39	.34	.43	.38	.33	.31
D2	.39	.40	.43	.44	.36	.43	.35	.38	.48	.45

Table 11
Average $\hat{b}$, Standard Deviation of $\hat{b}$, and Average Absolute Deviations Between $\hat{b}$ and d (AAD) for Item Set WT, From Original and Replication Samples, for Levels of the Correlation Between θ_1 and θ_2

	$r(\theta_1,\theta_2)$				
Sample and Statistic	.20	.40	.60	.80	1.00
Original Sample					
Average	.06	.03	.03	.01	.02
SD	1.62	1.49	1.39	1.32	1.25
AAD	.29	.18	.10	.05	.05
Replication 1					
Average	.00		.01		
AAD	.30		.12		
Replication 2					
Average	.05		.04		
AAD	.29		.12		
Replication 3					
Average	.05		.01		
AAD	.30		.13		
Replication 4					
Average	.02		−.03		
AAD	.28		.12		

References

Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9,* 37–48.

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

Bloxom, B., & Vale, C. D. (1987, June). *Multidimensional adaptive testing: An approximate procedure for updating*. Paper presented at the Psychometric Society meetings, Montreal.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179–197.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement, 7,* 189–199.

Green, B. F. (1988). *Notes on the test information function in the multidimensional compensatory IRT model* (Report 88-10). Baltimore: Johns Hopkins University, Psychometric Laboratory.

Hambleton, R. K., & Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10,* 287–302.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

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

McDonald, R. P. (1965). Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology, 18,* 11–23.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34,* 100–117.

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 6,* 379–396.

McKinley, R. L., & Reckase, M. D. (1983a). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. *Behavior Research Methods and Instrumentation, 15,* 389–390.

McKinley, R. L., & Reckase, M. D. (1983b). *An ex-*

Table 12
Average $\hat{a}$ and Average Absolute Deviations Between $\hat{a}$ and a_T for Each Subset and Dimension Group of Item Set WT, From Original Samples, for Levels of the Correlation Between θ_1 and θ_2

Statistic and Subset	$r(\theta_1,\theta_2)$ .20	.40	.60	.80	1.00
Average					
$\hat{a}$(S1)	.91	1.03	1.17	1.44	1.81
$\hat{a}$(S2)	.98	1.12	1.31	1.38	1.59
$\hat{a}$(S3)	1.20	1.25	1.44	1.57	1.71
$\hat{a}$(S4)	1.31	1.47	1.42	1.52	1.61
$\hat{a}$(S5)	1.19	1.30	1.44	1.51	1.61
$\hat{a}$(S6)	1.09	1.39	1.39	1.39	1.63
$\hat{a}$(S7)	.92	1.03	1.19	1.40	1.59
AAD					
S1	.63	.52	.38	.19	.31
S2	.48	.33	.20	.13	.16
S3	.31	.25	.16	.18	.23
S4	.24	.19	.17	.12	.14
S5	.26	.18	.12	.11	.16
S6	.41	.44	.16	.13	.18
S7	.62	.51	.35	.18	.14
Average					
$\hat{a}$(D1)	1.03	1.13	1.30	1.47	1.71
$\hat{a}$(D2)	1.07	1.24	1.34	1.43	1.61
AAD					
D1	.47	.37	.24	.17	.23
D2	.43	.38	.21	.14	.16
|AAD(D1)−AAD(D2)|	.04	.01	.03	.03	.08

Table 13
Average $\hat{a}$ and Average Absolute Deviations Between $\hat{a}$ and a_T for each Dimension Group of Item Set WT, From Original (1) and Replication (2-5) Samples, for Correlation Between θ_1 and θ_2 of .20 and .60

	$r(\theta_1,\theta_2) = .20$					$r(\theta_1,\theta_2) = .60$				
Statistic	1	2	3	4	5	1	2	3	4	5
Average										
a(D1)	1.03	1.04	1.00	1.01	1.07	1.30	1.32	1.34	1.26	1.25
a(D2)	1.07	1.05	1.08	1.10	1.05	1.34	1.32	1.33	1.32	1.31
AAD										
D1	.47	.49	.50	.49	.43	.24	.23	.30	.27	.30
D2	.43	.45	.42	.41	.45	.21	.32	.23	.20	.21
|AAD(D1)−AAD(D2)|	.04	.03	.08	.08	.02	.03	.09	.07	.07	.09

tension of the two-parameter logistic model to the multidimensional latent space (Research Report ONR83-2). Iowa City IA: American College Testing Program.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177–195.

Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association, 70,* 351–356.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207–230.

Reckase, M. D. (1985a, April). *The difficulty of test items that measure more than one ability*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Reckase, M. D. (1985b, August). *Trait estimates from multidimensional items*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles.

Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement, 25,* 193–203.

Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika, 39,* 111–121.

Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika, 51,* 589–601.

Sympson, J. B. (1978). A model for testing with multidimensional items. In D. J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference*. Minneapolis: University of Minnesota.

Thomas, T. J., & Green, B. F. (1989). *Item presentation controls for computerized adaptive testing: Content-balancing vs. mini-CAT* (Report 89-1). Baltimore: Johns Hopkins University, Psychometric Laboratory.

Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimation. *Applied Psychological Measurement, 12,* 239–252.

Wingersky, M. S. (1984). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *Applications of item response theory*. Vancouver: Educational Research Institute of British Columbia.

Acknowledgments

This research was supported, in part, through funding from the Office of Naval Research, Contract No. N 00014-86-K-0653. The authors thank Jeffrey J. Walczyk for assistance in running simulations and statistical analyses.

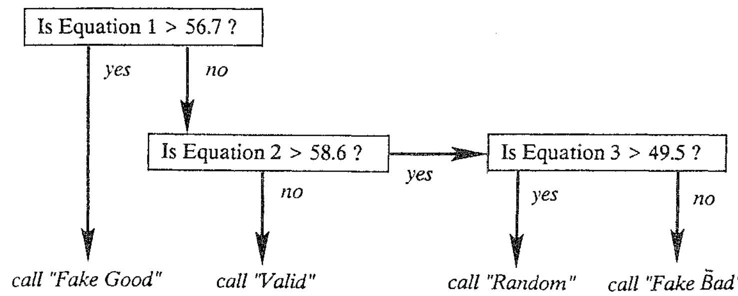
Author's Address

Send requests for reprints or further information to Valerie Greaud Folk, Educational Testing Service, Princeton NJ 08541, U.S.A.

Correction

Lanning, K. (1989). Detection of invalid response patterns on the California Psychological Inventory. Volume 13, Number 1, pp. 45–56.

Figure 1 on page 48, "Decision Tree for Evaluating CPI Profile Validity," requires two corrections: The labels for the branches from the Equation 2 node should be reversed, and the "Valid" and "Fake Good" outcomes should be exchanged. The corrected figure follows:



Announcement Correction

The Hebrew University of Jerusalem Louis Guttman Memorial Fund announcement in the March, June, and September 1989 issues of *Applied Psychological Measurement* gave an incorrect address for donations within the United States. The correct address is **American Friends of the Hebrew University, 11 East 69th Street, New York NY 10021.** Checks from Americans should be made out to American Friends of the Hebrew University—Louis Guttman Memorial Fund. As before, gifts from outside the U.S. should be sent to The Hebrew University of Jerusalem, Division for Development and Public Relations, 91905 Jerusalem, Israel. Donations are tax-deductible in most countries.