MODELING INCORRECT RESPONSES TO MULTIPLE-CHOICE ITEMS
WITH MULTILINEAR FD (U) ILLINOIS UNIV AT URBANA MODEL
BASED MEASUREMENT LAB F DRASGOW ET AL AUG 87
UNCLASSIFIED MEASUREMENT SER-87-1 N00014-83-K-0397 F/G 5/8 ML
MODELING INCORRECT RESPONSES TO MULTIPLE-CHOICE ITEMS WITH
MULTILINEAR FORMULA SCORE THEORY

Fritz Brenguier, Michael V. Levine, Bruce Williams,
Mary E. McLaughlin, and Gregory Candell

Rand-based Measurement Laboratory
University of Illinois
210 Eddyson Building
Champaign, IL 61820

August 1987

Prepared under contract No. N00014-83-K-0397, NR 150-518
and No. N00014-85-K-0468, NR 421546
with the Personnel and Training Research Program
Psychological Sciences Division
Office of Naval Research

Approved for public release: distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.
Modeling Incorrect Responses to Multiple-Choice Items with Multilinear Formula Score Theory

Multilinear Formula Score theory provides powerful methods for solving psychological measurement problems of long standing. In this paper the question of information in incorrect option selection on multiple choice items is addressed. Multilinear formula scoring (MFS) is first used to estimate option characteristic curves for the Armed Services Vocational Aptitude Battery Arithmetic Reasoning test. Accurately estimated curves are obtained for real and simulated data. Then the statistical information about ability is computed for dichotomous and polychotomous scorings of the items. Moderate gains in information are obtained for low to slightly above average abilities. The dichotomous and polychotomous models are then compared for their relative performances in appropriateness measurement. The rates of detection of some types of aberrance responding were more than 100% higher for optimal polychotomous model index. Consequently the MFS polychotomous model provides opportunities for better testing by allowing more accurate ability estimates, improvements in the theory and practice of item writing, and more powerful appropriateness measurement.
18. Subject terms (continued)

appropriateness measurement, optimal statistical test, cheating, coaching.
Abstract

Multilinear Formula Score theory provides powerful methods for solving psychological measurement problems of long standing. In this paper the question of information in incorrect option selection on multiple choice items is addressed. Multilinear formula scoring (MFS) is first used to estimate option characteristic curves for the Armed Services Vocational Aptitude Battery Arithmetic Reasoning test. Accurately estimated curves are obtained for real and simulated data. Then the statistical information about ability is computed for dichotomous and polychotomous scorings of the items. Moderate gains in information are obtained for low to slightly above average abilities. The dichotomous and polychotomous models are then compared for their relative performances in appropriateness measurement. The rates of detection of some types of aberrance responding were more than 100% higher for optimal polychotomous appropriateness indices than any dichotomous model index. Consequently the MFS polychotomous model provides opportunities for better testing by allowing more accurate ability estimates, improvements in the theory and practice of item writing, and more powerful appropriateness measurement.
Introduction

Multilinear formula score theory or multilinear formula scoring (MFS; Levine, 1985a, 1985b) is a nonparametric item response theory for which consistent and asymptotically efficient estimators of ability densities, item characteristic curves (ICCs), and option characteristic curves (OCCs) have been derived and programmed. MFS provides a powerful new approach to substantive questions of long standing. These questions include determining the shapes of ability distributions and the magnitudes of differences between ability distributions of various groups, determining the shapes of item characteristic curves for unidimensional and multidimensional tests, identifying biased and other faulty items, and assessing the extent to which two tests measure the same ability.

In this paper we focus on MFS's ability to estimate efficiently option response curves from small samples for responses that may fail to satisfy the local independence assumption of item response theory. The benefits of this endeavor shall be assessed in two ways. First, we determine the increase in information about ability due to polychotomous scoring of item responses. Here the term "information" is used in its statistical sense to mean the expected squared derivative of the logarithm of the likelihood function. Since the asymptotic standard error of the maximum likelihood estimate of an ability θ equals the square root of the reciprocal of the information function at θ, an increase in information due to polychotomous scoring is readily translated into percent test length reduction made possible by polychotomous scoring.

The second comparison is between the dichotomous and polychotomous item response model's potentials for supporting appropriateness measurement. Levine and Rubin (1979) introduced this term to refer to model-based methods
for detecting response patterns that yield faulty measures of ability. For example, test scores are spuriously high when a low ability examinee copies some answers from a high ability neighbor or has been given answers to some questions by an informant. Spuriously low test scores result from alignment errors, atypical educations, unusual creativity, deliberate failure, and a variety of other sources.

Of course, the model-based detectability of a particular type of aberrance depends upon the item response model used to analyze the data; more specific (polychotomous) models are expected to be rejected more frequently when fitted to aberrant response patterns and thus provide superior appropriateness measurement. Recently Levine and Drasgow (1984, 1987) developed a technique for computing the power of the most powerful appropriateness measurement procedure supported by an item response model. By combining the new optimality results with MFS's ability to accurately recover the option characteristic curves needed for polychotomous modeling we intend to determine whether polychotomous modeling is negligibly or markedly superior to dichotomous modeling in detecting test anomalies.

This study also contributes to formula score theory in that it contains a verification of MFS theoretical results with simulation data.

Review of Multilinear Formula Score Theory

In this section we review MFS theory as it is used in this paper. The theory is more general than outlined here, but for the sake of clarity we describe only the special case required for the present research.

Let \( u_i \) denote the response to the \( i \)th item of an \( n \) item test scored \( u_i = 1 \) if correct and \( u_i = 0 \) if incorrect. The \( u_i \) generate the elementary formula scores, which can be enumerated as
Traditional formula scoring (Lord and Novick, 1968, especially Chapter 14) generally uses only linear scores. When there is neither omitting nor polychotomous scoring, **linear formula scores** are formulas with a constant term plus a linear combination of the binary item scores, $u_1, u_2, ..., u_n$. (When there is omitting and polychotomous scoring, a linear score is a constant plus a linear combination of binary variables indicating omitting and option choice.)

Multilinear formula score theory generalizes traditional formula score theory by using quadratic scores (linear scores added to linear combinations of $u_1u_2, u_1u_3, ..., u_{n-1}u_n$), cubic scores (quadratic scores plus linear combinations of products of item scores for three different items), and higher order scores. Most of the results in this paper were obtained with fifth order scores. The new theory is called "multilinear" because frequent use is made of the fact that when all the scores except one are held constant, a "linear" score is obtained.

In this paper we shall assume that the regression of $u_1$ on the latent trait $\theta$ is a three-parameter logistic ogive

$$E(u_1 \mid \theta = t) = c_1 + \frac{1 - c_1}{1 + \exp[-Da_1(t - b_1)]} = P_1(t),$$
where $D$ is a scaling constant set equal to 1.702, $a_i$ is the discrimination parameter, $b_i$ is the difficulty parameter, and $c_i$ is the lower asymptote of the ICC. By local independence, the regressions of the elementary formula scores on the latent trait can then be written

$$
\begin{align*}
1 \\
P_1(t), P_2(t), \ldots, P_n(t) \\
P_1(t)P_2(t), P_1(t)P_3(t), \ldots, P_{n-1}(t)P_n(t) \\
\vdots \\
P_1(t)P_2(t) \ldots P_n(t),
\end{align*}
$$

where each $P_i(t)$ is a three-parameter logistic ICC.

There are $2^n$ regression functions listed above. More can be generated by taking linear combinations of the elementary formula scores and then computing their regressions on the latent trait. For example, the number-right score

$$
X = u_1 + u_2 + \ldots + u_n,
$$

has the regression

$$
E(X \mid t) = \sum_{i=1}^{n} P_i(t).
$$

The collection of regression functions of all linear combinations of elementary formula scores is called the **canonical space** of a test.

A major step in a MFS analysis of a test consists of finding a smaller number of functions to represent the large number (in fact, an infinite number) of functions in the canonical space. The smaller collection of functions is called an **orthonormal basis** for the canonical space.
Selecting an orthonormal basis for the canonical space is analogous to finding the principal components of a set of variables. In a principal components analysis, the basic idea is to create a new set of variables, the principal components, so that each of the original variables can be written as a linear combination of the principal components plus a small residual. A principal components analysis is valuable when there is a large number of original variables and the first few principal components explain almost all of their variance. In the same way functions in the canonical space are written as linear combinations of the orthonormal basis functions. For example, the ICC for the ith item can be written

\[ p_i(t) = \sum_{k=1}^{K} a_k h_k(t), \]

where \( K \) functions, denoted \( h_1(t), \ldots, h_K(t) \) are used in the orthonormal basis and the \( a_k \) are the weights used in the linear combination. If \( K \) is sufficiently large, this representation is exact. If only the first \( J \) functions are used, instead of all \( K \) functions (where \( J \) is less than \( K \)), then there is some error. However, the residual

\[ p_i(t) - \sum_{k=1}^{J} a_k h_k(t) = \sum_{k=J+1}^{K} a_k h_k(t) \]

will be small if the \( a_k \) are small for values of \( k \) larger than \( J \). In fact, the area under the squared residual is exactly

\[ a_{J+1}^2 + a_{J+2}^2 + \ldots + a_K^2. \]

In each MFS analysis a parsimonious representation of one or another collection of functions in the CS is important. MFS provides techniques that yield basis functions that give small \( a_k \) for large \( k \), at least for the collection of functions being analyzed. Most MFS analyses require six
to eight basis functions for an adequate representation of the functions being studied. Ten were used in this study.

To recapitulate, the analysis begins by estimating ICCs from the dichotomously scored item responses. Widely available programs such as LOGIST and BILOG can be used to this end. The estimated ICCs (and the assumption of local independence) are subsequently used to define the canonical space. Then a small number of orthonormal basis functions are selected so that the functions in the canonical space are well-approximated by linear combinations of the orthonormal basis functions.

The next step of the MFS analysis is to use the orthonormal basis functions to represent the option characteristic curves (OCCs). For technical reasons (see below), we first estimate orthonormal basis function weights for conditional option characteristic curves (COCCs). A COCC gives the probability of an option choice given that the person does not choose the correct option. A COCC equals its associated OCC divided by \((1-P_i(\theta))\). Hence the COCCs for an item sum to 1 for all \(\theta\) values whereas the OCCs sum to \(1-P_i(\theta)\), which becomes very small for large \(\theta\) values. Each option characteristic curve is then represented as the product of two linear combinations of the \(h_j\)'s, namely the representation of \(1-P_i\) and a COCC.

At this point the OCC can be represented by a single set of weights by calculating weights \(b_j\)'s such that \(\sum b_j h_j(\cdot)\) is approximately equal to \((1-P_i)\) times the COCC value. (An exact representation is not possible in general because a product of two functions in the canonical space is not necessarily in the canonical space.)

Since OCCs and COCCs were not included in the set of functions used to define the canonical space, there is both the mathematical question of how best to approximate the OCCs and COCCs by basis functions and the
substantive question of whether or not the basis functions can adequately approximate OCCs and COCCs. The analysis proceeds item-by-item with the weights for all the options (including omit as an option) to each item simultaneously estimated by "marginal" maximum likelihood. The log likelihood that is maximized with respect to the weights is

\[
L = \sum_{j=1}^{N} \log P(u_j^*, v_{ij}^*) ,
\]

where \( u_j^* \) is a vector containing the dichotomously scored item responses of the \( j \)th examinee and \( v_{ij}^* \) indicates the particular option on item \( i \) selected by examinee \( j \). For a four option multiple-choice item, \( v_{ij}^* = 1 \) if option A is selected, \( ... v_{ij}^* = 4 \) if option D is selected, and \( v_{ij}^* = 5 \) if no response is made. Suppose all the items are recoded so that option A is always the correct response. Then Equation 1 can be rewritten as

\[
L = \sum_{j=1}^{N} \log P(u_j^*) + \sum_{v_{ij}^* = 1}^{5} \log \int P(u_j^* | t) P(v_{ij}^* | t, u_{ij} = 0) f(t) dt
\]

where

\[
P(u_j^* | t) = \prod_{i=1}^{n} P_i(t)^{u_{ij}} (1 - P_i(t))^{1-u_{ij}} ,
\]

\[
P(v_{ij}^* | t, u_{ij} = 0) = \sum_{k=1}^{J} \alpha_k h_k(t) ,
\]
and \( f(t) \) is the ability density. Notice that Equation 3 is the likelihood function for the three-parameter logistic model (i.e., Lord's (1980) Equation (4-20) and Hulin, Drasgow, & Parson's (1983) Equation (2.6.2)). It is the \( \alpha_k \)'s in Equation 4 that are to be estimated. Actually, each option has its own set of \( J \alpha_k \)'s, but to avoid notational complexity we have not added another subscript to the \( \alpha_k \)'s.

It is important to observe that local independence is not used to derive Equation 2 from Equation 1; only the definition of conditional probability is used. Thus, even when skipping items or not teaching items (response "5") fails to obey the assumption of local independence, an accurate estimate of the conditional probability of non-response for examinees at each ability level may be obtained.

Quadratic programming is used to obtain maximum likelihood estimates of orthonormal basis function weights for the COCCs in Equation 4. The weights \( \alpha_k \) for the COCCs are easier to estimate than the weights for OCCs since the OCCs for easy items and OCCs for rarely chosen options are close to zero, which causes the \( \alpha_k \) to become indeterminate; COCCs are not usually close to zero. Since the OCC at \( \theta = t \) is equal to the COCC times \( 1 - P_i(t) \), the OCCs are available after the COCCs have been obtained. The COCCs are intrinsically interesting as well as mathematically tractable since their shapes can be used to study the properties of effective distractors.

The quadratic programming methods used by Levine and Williams (1987) are convenient because they allow plausible constraints to be placed on the COCCs. One constraint is positivity: COCCs are not allowed to become negative. In our analyses all COCCs were required to equal or exceed .001. A second constraint placed on COCCs is smoothness: The COCCs were not allowed to oscillate widely. The smoothness constraint can be implemented.
by restricting the third derivative of the COCCs to be less than .005. This condition can be thought of as requiring each small piece of the graph of the COCC to have a very accurate quadratic approximation. (A restriction on the second derivative would force the COCC to be locally linear and a first derivative constraint would force the COCC to be locally constant.)

In summary, orthogonal basis functions \( h_j(t) \) are derived from IICs, which are estimated by programs such as LOGIST or BILOG. COCCs are represented as linear combinations of the basis functions in Eq. 4, and marginal maximum likelihood estimates of the weights \( a_j \) in this equation are obtained. OCC values can then be obtained by multiplying COCC values times \((1-P_i)\).

**Estimation and Information**

**Data set.** The data set used in our analyses was a spaced sample of 2978 examinees; this data set is fully described in the Profile of American Youth (1982). These examinees answered the 30 item Armed Services Vocational Aptitude Battery (ASVAB) Arithmetic Reasoning (AR) test. Each item on this test has four options.

**ICC estimation.** The first step in the MFS analysis is to estimate ICCs from the dichotomously scored item responses. To this end, the item responses of the examinees described above were scored dichotomously. All unanswered items were scored as incorrect (since skipping and not reaching are treated as a separate—and incorrect—response option). Then the LOGISTIC (version 2B) computer program (Wood, Wingersky, & Lord, 1976) was used to estimate item and person parameters. Estimates of item discrimination parameters ranged from about 0.5 to 2.0 and estimates of item difficulties varied from about -3.0 to 1.4 (mean = .14, SD = .99).
NFS Theory

Informative, suggesting a relatively large number of examinees with very low abilities. The substantive interpretation of this fat left tail is that most examinees answered more than half of the items there may have been some who were poorly motivated and did not have a serious attempt to answer the test. This was confirmed by the examination of the data for the test, indicating that the motivation was adequate for adequate motivation. The test items were not of sufficient variance and consequently simultaneously examined the total three examinees.

A serious concern should be

...
for all 11,914 examinees in the American Youth data set, forming 25 ability strata on the basis of estimated abilities by using the 4th, 8th, ..., 96th percentile points of the standard normal distribution as cutting scores, and then computing the proportion of examinees selecting each option among the subset of examinees who answered the item incorrectly. The centers of the vertical lines correspond to the observed proportions and they are plotted above the category medians (the 2nd, 6th, ..., 98th percentile points of the standard normal distribution). The vertical lines represent approximate 95% confidence intervals for the observed proportions (± two standard errors, where the observed proportion is used to compute the standard error). Observed proportions of 0 and 1 are plotted as plus signs and are offset slightly from their true locations so that they will be visible.

The AR items seem to be more-or-less ordered by difficulty. Consequently, the 95% confidence intervals for the first few items in Appendix I are very wide because these items are easy and so few examinees choose incorrect options. Confidence intervals for later items are much narrower and provide a severe test for COCC estimates. Item 27, for example, shows that the COCC estimates provide a very good description of option choice. Notice that the COCC for the omit category lies below most observed proportions. This occurs because examinees with high omitting rates were excluded from the sample used to estimate COCCs, but were included in the total sample used to compute the proportions displayed in Appendix I.

COCC estimation verification. The figures presented in Appendix I show that MFS estimates of COCCs closely follow the actual patterns of item responses. It is difficult, however, to understand the accuracy of COCC estimates from these figures because the true COCCs are not known. To gain
further insights into the properties of MFS estimates of COCCs, a simulation data set of 3000 response patterns was generated. Simulated abilities were sampled from the standard normal distribution, probabilities of correct and incorrect responses were determined from the ICCs obtained by the LOGIST run described previously, and probabilities of option selections (for responses simulated to be incorrect) were computed using the MFS estimated COCCs. Thus, the assumptions used to estimate COCCs correspond exactly to the way in which the data set was generated.

COCCs were re-estimated from the simulation data set. The true ability density (the standard normal) was used in Equation 2 and the true ICC values were used to compute probabilities of correct and incorrect responses. The true ability density and ICC values were used because we wanted to determine the errors of COCC estimates in a way that was not confounded with inaccuracies in density estimates and ICC estimates.

The results of the simulation study are shown in Appendix 2, which presents the re-estimated COCCs for all 30 items. Heavy lines indicate the re-estimated COCCs and thin lines indicate the true COCCs. Observed proportions and their approximate 95% confidence intervals are shown for the simulation sample of N = 3000. The observed proportions are not plotted if five or fewer incorrect responses were made in an ability stratum.

Item 2 shows estimated COCCs that are very close to the true COCCs for all ability levels. This is remarkable because there were almost no incorrect responses made by simulated examinees with above average ability. Item 3 shows that we cannot always expect to have well-estimated COCCs when there are no data available: Large differences between true and estimated COCCs occur at high ability levels. The COCCs are, however, accurately
estimated in ability ranges for which there were very few incorrect
responses.

From an inspection of the plots in Appendix 2 it seems evident that
COC values are accurately estimated when there are six or more incorrect
responses in adjacent ability strata. Sometimes COC values are well-
estimated when fewer incorrect responses are available, but this seems to be
a matter of chance. Notice, also, that COCs for the omit option are not
underestimated in this analysis as they were in the analysis of the real AR
data. In the present analysis, all response vectors were used; there was no
restriction on omitting as in the previous analysis. In this simulation
study data were unidimensional in the sense that the probability of omitting
depended only on ability, although it was permitted to vary from item to
item. It would have been more realistic to use a two dimensional simulation
model with examinees varying both in ability and tendency to omit.

**Information function.** Information functions for the dichotomous and
polychotomous modelings of the AR test are shown in Figure 2. An expression
for the information function of the three-parameter logistic model is

\[
\text{Information at } t = \frac{[P_i'(t)]^2}{P_i(t) Q_i(t)} + \frac{[Q_i'(t)]^2}{Q_i(t)}
\]

(5)

where \( Q_i = 1 - P_i \) and \( P_i' \) and \( Q_i' \) are the first derivatives of \( P_i \) and
\( Q_i \). The information function of the polychotomous model is

\[
\text{Information at } t = \frac{[P_{ij}'(t)]^2}{P_{ij}(t) Q_{ij}(t)} + \frac{[Q_{ij}'(t)]^2}{Q_{ij}(t)}
\]

(6)

where \( P_{ij} \) is the OCC for option \( j \) on item \( i \) and \( P_{ij}' \) is its first
derivative. The correct option makes the same contribution to information
for both the dichotomous and polychotomous scorings, namely, the first term.
on the right sides of Equations 5 and 6. Thus, any differences in
information are entirely due to the treatment of incorrect responses. Using
Jensen's inequality (Halmos, 1950) it can be shown that

\[
\sum_{j=2}^{J} \frac{[P_{ij}(t)]^2}{P_{ij}(t)} > \frac{[Q_i(t)]^2}{Q_i(t)}
\]

(cf., Samejima, 1969; Park, 1983). Thus, any increase in information is
entirely due to polychotomous scoring.

---

Figure 2 shows that there are moderate gains in information due to
polychotomous scoring of the AR items for low to moderately high abilities.
Little or no information is gained for high ability examinees; this latter
finding is not surprising because high ability examinees are expected to
answer nearly all the items correctly.

It should be noted that the AR items were not written with
polychotomous scoring in mind and so the gains in information shown in
Figure 2 are more-or-less accidental. Larger gains might be realized if
item writers knew the attributes of incorrect options that typically lead to
substantial increases in information.
Appropriateness Measurement

Purpose

In this section we compare the effectivenesses of dichotomous and polychotomous models for detecting aberrant responses patterns. By comparing detection rates of optimal indices it is possible to compare the maximum detection rates possible for a given form of aberrance. In this section, as in the previous section, the dichotomous model is a submodel of the polychotomous model; hence any increase in detection rates is due to modeling incorrect responses.

For an optimal index to be truly optimal, it must be computed from the true ICCs or OCCs and, therefore, the optimal indices for dichotomous and polychotomous scorings of the simulation data were computed using the simulation ICCs and OCCs. In any practical application, however, only estimated ICCs and OCCs will be available. Consequently, we decided to examine one aspect of the robustness of optimal indices by computing the optimal index for dichotomously scored responses using ICCs estimated by the LOGIST (Wood, Wingersky, & Lord, 1976) computer program. Further research designed to develop extensions of optimal indices for use in practical settings will be warranted if the optimal indices computed from estimated ICCs are found to be nearly as powerful as optimal indices computed from the true ICCs.

Several practical indices were also evaluated. Most of these indices were computed from the dichotomously scored item responses. One index, however, is the natural extension of a dichotomous model index to the polychotomous case. Detection rates for the practical indices indicated (1) which were relatively more powerful and less powerful; and (2) the extent to which the maximum detection rates were attained.
Overview

The ICCs and OCCs estimated for the AR test from the sample of N = 2,978 were used as the "true" item parameters in a simulation study. Initially, a sample of N = 3000 simulated response patterns was created and used as a test norming sample. This data set was used to determine the item and test statistics required to compute all but one (zp) of the practical appropriateness indices listed in the next section. Then a normal sample of N = 4000 responses vectors was created. In addition, sixteen aberrant samples of N = 2000 were generated to simulate several forms of aberrance. Optimal indices and all the practical indices were then computed for the normal sample and aberrant samples. Rates of detection of aberrant responses vectors at various false alarm rates were determined for each appropriateness index and each form of aberrance.

Appropriateness Indices

In this section we list the appropriateness indices that are evaluated. For the sake of brevity we shall not provide extensive technical detail. This information is given by Levine and Drasgow (1984; 1987) for optimal indices and by Drasgow, Levine, and McLaughlin (1987) for practical indices. Additional references are given when appropriate.

Polychotomous model optimal indices (LRp). Levine and Drasgow (1984) used the Neyman-Pearson lemma to derive a class of most powerful appropriateness indices. These indices require the probabilities of observing the polychotomously scored response vector v* assuming that it was generated by a normal process (PNormal (v*)) and assuming that it was generated by a specified aberrant process (PAberrant (v*)). The decision procedure that classifies response vectors as aberrant when

PAberrant (v*) \leq \text{constant} \cdot P_{\text{Normal}} (v*) ,
where the constant is chosen to control the false alarm rate or Type I error rate, is at least as powerful as any other test. Thus, the polychotomous model optimal indices studied here have the form

\[ LR_p = \frac{P_{Aberrant}(v^*)}{P_{Normal}(v^*)}, \]

where the probabilities are computed using three-parameter logistic ICCs to determine conditional probabilities of correct responses and MFS OCCs to determine conditional probabilities of incorrect responses. Technical details about the form of \( LR_p \) for specific types of aberrance and an efficient computing algorithm are given by Levine and Drasgow (1984; 1987).

**Dichotomous model optimal indices (LR).** These indices are identical to the \( LR_p \) indices except that the only information used in their calculation is the pattern of correct and incorrect responses \( u^* \), i.e., the dichotomously scored item responses. This class of indices, therefore, provides the highest rates of detection when the choice of incorrect option is ignored.

**Dichotomous model optimal indices computed using estimated item parameters (LR̂).** For optimal indices to be truly optimal they must be computed using item parameters -- not item parameter estimates. In previous work (Levine & Drasgow, 1982), we found that the values of some appropriateness indices were almost unaffected when item parameter estimates were used in place of item parameters. In the present research, optimal indices for the three-parameter logistic model were also computed using estimated item parameters.

**Dichotomous and polychotomous model standardized \( l_0 (z, and z_p).** These indices, originally developed by Drasgow, Levine, and E. Williams (1985), are well-standardized (i.e., their conditional distributions given
ability are nearly invariant across ability levels) and are, therefore, well-suited to practical applications. In essence, they compare the likelihood of a vector of item responses to the expected likelihood given the examinee’s ability estimate. In previous research (Levine & Rubin, 1979; Levine & Drasgow, 1982; Drasgow, Levine, & E. Williams, 1985), it has been found that aberrant response vector tend to have likelihoods that are smaller than expected of normal response vectors, and thus, the standardized likelihoods $z$ and $z_p$ serve as effective appropriateness indices.

**Fit statistic (F1 and F2).** Two fit statistics suggested by Rudner (1983) as generalizations of Rasch model fit statistics used by Wright and his colleagues are

$$F1 = \frac{1}{n} \sum_{i=1}^{n} \frac{[u_i - P_i(\hat{\theta})]^2}{[P_i(\hat{\theta})Q_i(\hat{\theta})]}$$

and

$$F2 = \frac{1}{n} \sum_{i=1}^{n} \frac{[u_i - P_i(\hat{\theta})]^2}{\sum_{i=1}^{n} P_i(\hat{\theta})Q_i(\hat{\theta})}.$$  

Notice that $F1$ and $F2$ tend to be large when an examinee misses items ($u_i = 0$) that should be answered correctly ($P_i(\hat{\theta})$ near 1) and correctly answers ($u_i = 1$) items that should be very difficult ($P_i(\hat{\theta})$ near 0). Drasgow, Levine, and McLaughlin (1987) found $F2$ to be well-standardized. $F1$, however, was badly standardized because relatively many large values were observed for simulated normal, high ability examinees.

**Caution indices (S, T2, and T4).** Three "caution" indices were evaluated. The first is the original Sato caution index $S$ described by Sato (1975) and Tatsuoka and Linn (1983). The other two caution indices are the second (T2) and fourth (T4) standardized extended caution indices.
developed by Tatsuoka (1984). Drasgow, Levine, and McLaughlin (1987) found T4 to be better standardized than T2, so T4 should be preferred when their detection rates are comparable.

**Likelihood function curvature statistics** (JK and O/E). It is expected that the likelihood function will be "flatter" for aberrant response vectors than normal response vectors at the maximum likelihood ability estimate \( \hat{\theta} \). Two indices that provide measures of the flatness of the likelihood function were evaluated. The first (JK) is a normalized jackknife estimate of the variance of \( \hat{\theta} \) and the second is the ratio of the observed and expected information about ability contained in the dichotomously scored item responses. Both of these indices are described by Drasgow, Levine, and McLaughlin (1987), who showed that JK and O/E are well standardized.

**Method**

**Data Sets.** A test norming sample of 3000 responses vectors was created by sampling 3000 numbers (0's) from the normal (0,1) distribution truncated to the [-5.0, 3.5] interval. A normal sample of 4000 response vectors was also generated in this way. Two thousand aberrant response vectors were created in each of sixteen conditions. These conditions resulted from varying three factors: the type of aberrance (spuriously high; spuriously low), the severity of aberrance (mild; moderate), and the distribution from which simulated abilities were sampled.

Eight of the aberrant samples contained spuriously high response vectors and the remaining eight samples contained spuriously low responses vectors. Spuriously high responses patterns were created by first generating normal response vectors (using the AR three-parameter logistic ICCs to determine the probabilities of correct responses and the AR COCCs to determine probabilities of incorrect option selection) and then replacing a
given percentage $k$ of simulated responses (randomly sampled without replacement) with correct responses. Spuriously low response patterns were also created by first generating normal response vectors. Then a fixed percentage of items were randomly selected without replacement and the responses to these items replaced with random responses (i.e., a response was replaced by option A with probability .25, by option B with probability .25, ..., and by option D with probability .25). Mildly aberrant response patterns were generated by using $k = 17\%$ (i.e., 5 of 30 items). Moderately aberrant response patterns were created using $k = 33\%$ (i.e., 10 of 30 items).

The third variable manipulated was the ability level of the aberrant sample. Abilities for the spuriously high samples were sampled from four parts of the normal (0,1) distribution truncated to $[-5.0, 3.5]$: very low (0th through 9th percentiles), low (10th through 30th percentiles), low average (31st through 48th percentiles), and high average (49th to 64th percentiles). In all cases percentile points were determined after the truncation to $[-5.0, 3.5]$. These intervals were used because it is more important to detect spuriously high response patterns for low ability examinees than for high ability examinees. Similarly, it is more important to detect spuriously low responses by high ability examinees. Consequently, abilities were sampled from four average to high ability strata for the spuriously low samples: very high (93rd percentile and above), high (65th through 92nd percentiles), high average (49th through 64th percentiles), and low average (31st to 48th percentiles). The ability percentiles used here correspond to the percentiles forming United States Armed Service Vocational Aptitude Battery mental categories.
Analysis. All the item and test statistics required to compute the practical appropriateness indices were computed using the test norming sample. These quantities were computed as the first step in the analysis and then used in all subsequent analyses. LOGIST (Wood, Wingersky, & Lord, 1976) was used to estimate three-parameter logistic item parameters and a Fortran program was written to compute the other quantities required.

The practical appropriateness indices and \( LR_p \) were then computed for the 4000 response vectors in the normal sample. The item and test statistics estimated from the test norming sample were used in these calculations. This procedure simulates the process by which practical appropriateness indices would be computed in many applications. Optimal indices were also computed for the normal sample for four aberrant conditions: 17% spuriously high, 33% spuriously high, 17% spuriously low, and 33% spuriously low. The ICCs and COCCs used to generated the data were used to compute \( LR_p \) and \( LR_s \).

The practical appropriateness indices were computed for each of the 16 aberrant samples. In addition, the 17% spuriously high optimal index was computed for the four samples with this form of aberrance, the 33% spuriously high optimal index was computed for the four samples with this form of aberrance, etc. The proper interpretation of the optimal indices computed in the present research is the following: They are optimal for the specified form of aberrance, say 17% spuriously high, in a population where the ability density is a truncated normal for both the normal and aberrant populations and a response vector is either normal or 17% spuriously high. The normal group does in fact have this ability distribution. By stratifying on a subinterval of \([-5.0, 3.5]\) for the aberrant group, we
determined the power of the index that is optimal for the population as a whole in a particular subpopulation.

Evaluation Criteria. The main criteria for evaluating the appropriateness of indices were the proportions of aberrant response patterns correctly identified as aberrant when various proportions of normal response patterns were misclassified as aberrant. These proportions shall be presented for all 16 aberrance conditions. This allows us to determine what types of aberrant response patterns have acceptably high detection rates using optimal methods and using practical methods. The characteristics of response patterns that cannot be detected are evident as a consequence of examining the 16 aberrance conditions separately.

Results

The results for the spuriously high conditions are given in Tables 1 through 4. The results for the lowest ability group are shown in Table 1. In this table it is evident that cheating on five randomly selected items is not very detectable: At a 2% false alarm rate only 28% of the simulated cheaters are detected by the optimal $L_{DP}$ index. The best of the practical indices, $z$, and $F_2$, detect 18% and 20%, respectively. Cheating on 10 items (the 33% condition) is reasonably detectable. For example, $L_{DP}$ detects 61% and $L_R$ detects 54% at a 2% false alarm rate. At this false alarm rate, $z$, $F_2$, and $T_4$ detect 44%, 41%, and 50%, respectively. Finally, detection rates for optimal indices computed from true and estimated ICCs are very similar for almost all false alarm rates in Table 1.

Insert Tables 1 through 4 about here
The results further support the hypothesis that a $28$ expression is significantly greater than the $21$ expression. The average response of $32$ subjects was $28$.

The correlation coefficient is $r = 0.7$.

The null hypothesis is rejected at the $0.05$ level of significance.

The Pearson correlation coefficient is $r = 0.8$.

The regression line is $y = 2x + 1$.

The standard deviation is $s = 4$.

The confidence interval is $(0.75, 0.95)$.

The t-test statistic is $t = 2.3$.
A, has detection rates that are close to .85, in all of the tables. This entire power for detecting inappropriate response patterns is lost when the true three parameter logistic curves are replaced by estimated ones.

Incorporate tables through "about here.

Discussion

In this paper we have described and implemented Levine's (1985a, 1985b) approach to polychotomous measurement. It was used to estimate 11 I P for a sample of 250 examinees who responded to the ADVAM test. Good to excellent fits were obtained when the estimated curves were compared to empirical proportions computed from the responses of a larger sample of 750 examinees. A simulation data set was also used to investigate 1000 estimates. Very accurate estimates were obtained for ability ranges with sufficient numbers of examinees who responded incorrectly.

The test information function of the polychotomous model was found to be moderately larger than the three-parameter logistic information function for the 10 moderately high ability levels. Since there is information in incorrect options it seems prudent to use it if items are expensive to write, the number of items that can be administered is severely limited, or very accurate ability estimates are required. Furthermore, we can now study systematic and the differences in items with informative incorrect options and items with essentially noninformative incorrect options. It may be possible to identify different characteristics of these two types of items and thereby help item writers increase the information about ability.
provided by tests by writing items with highly informative incorrect options.

An appropriateness measurement simulation study was also conducted to compare the polychotomous model with a dichotomous submodel, namely the three-parameter logistic. Several important results were obtained. First, for the spuriously low treatment that simulates atypical educations, misgridimg answers to a portion of the test, unusual creativity, etc., we found that optimal three-parameter logistic appropriateness indices fell far short of their optimal polychotomous model counterparts. At some false alarm rates, the rates of detection of aberrant response vectors were more than 100% higher for the polychotomous optimal indices. Thus appropriateness measurement constitutes one important practical testing problem where substantial gains are made by the use of a polychotomous item response model.

The results of the appropriateness measurement simulation study also showed that the practical polychotomous model index \( z_p \) was not a particularly good index: Its detection rates were not close to optimal for either spuriously high or spuriously low treatments. This result, in conjunction with the results described, previously point to the need to devise better polychotomous appropriateness indices that can be used in practical situations.

A third result obtained in the appropriateness measurement research is that the \( z_2 \), \( F_2 \), and \( T_4 \) indices effectively detect aberrance in relation to three-parameter logistic optimal indices (but not polychotomous model optimal indices). Therefore, if one is satisfied with dichotomous scoring of item responses for some particular application, then \( z_2 \), \( F_2 \), \( T_4 \) can be used with confidence to detect inappropriate test scores.
Means for implementing appropriateness measurement in practical settings are discussed by Drasgow and Guertler (1987).

Finally, the LR,' indices provided detection rates that were nearly as high as the rates provided by the optimal LR, indices. Thus, the three-parameter logistic optimal indices seem to be robust to item parameter estimation error. This result is surprising because extensive computations are required to evaluate LR,'; small errors (in ICC values) would be expected to grow progressively larger as the computations progressed. Nonetheless, only small differences between values of LR, and LR,' were observed for individual response patterns. Thus, we are encouraged to continue research on "almost-optimal" indices that are based on likelihood ratios and could be used in practical settings.

Conclusion

COCO estimation provides opportunities to improve testing in a variety of ways: ability estimation, the theory and practice of item writing, appropriateness measurement. Applications in areas such as item and test bias and adaptive testing may also be fruitful. Consequently, we conclude that there is information in incorrect responses and that polytomous item response models can make important contributions to psychological testing.
References


Levine, M.V. (1985). Classifying and representing ability distributions. Measurement Series 85-1, Model-Based Measurement Laboratory, 210 Education Building, Department of Educational Psychology, University of Illinois, 1310 S. Sixth Street, Champaign, IL 61820. (b)


Levine, M.V., & Drasgow, F. (1984). Performance envelopes and optimal appropriateness measurement. Measurement Series 84-5, Model-Based Measurement Laboratory, 210 Education Building, Department of Educational Psychology, University of Illinois, 1310 S. Sixth Street, Champaign, IL 61820.


Acknowledgments

This research was supported by Contract No. N00014-83-K-0397, NR 150 518, from the Office of Naval Research, Michael V. Levine, Principal Investigator and by Contract No. F4168-84-D-0002 from the U.S. Air Force Human Resources Laboratory to the Human Factors and Logistics Division of Universal Energy Systems, Inc., 4401 Dayton-Xenia Road, Dayton, OH 45432. We wish to thank Malcolm Ree, Randy Park, and James Earles for their help.
<table>
<thead>
<tr>
<th>False Alarm Rate</th>
<th>LRp</th>
<th>LR2</th>
<th>LR3</th>
<th>z_P</th>
<th>z</th>
<th>F1</th>
<th>F2</th>
<th>S</th>
<th>T2</th>
<th>T4</th>
<th>JK</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>17% Spuriously High Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>04</td>
<td>04</td>
<td>01</td>
<td>00</td>
<td>03</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>03</td>
<td>06</td>
<td>00</td>
<td>08</td>
<td>00</td>
<td>04</td>
<td>04</td>
<td>02</td>
<td>02</td>
</tr>
<tr>
<td>.01</td>
<td>16</td>
<td>19</td>
<td>17</td>
<td>05</td>
<td>12</td>
<td>02</td>
<td>13</td>
<td>03</td>
<td>07</td>
<td>06</td>
<td>03</td>
<td>04</td>
</tr>
<tr>
<td>.02</td>
<td>28</td>
<td>29</td>
<td>26</td>
<td>08</td>
<td>18</td>
<td>04</td>
<td>20</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>06</td>
<td>07</td>
</tr>
<tr>
<td>.03</td>
<td>34</td>
<td>33</td>
<td>30</td>
<td>11</td>
<td>25</td>
<td>07</td>
<td>24</td>
<td>18</td>
<td>18</td>
<td>14</td>
<td>09</td>
<td>09</td>
</tr>
<tr>
<td>.04</td>
<td>38</td>
<td>37</td>
<td>34</td>
<td>13</td>
<td>29</td>
<td>10</td>
<td>28</td>
<td>24</td>
<td>22</td>
<td>18</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>.05</td>
<td>43</td>
<td>40</td>
<td>38</td>
<td>15</td>
<td>33</td>
<td>15</td>
<td>32</td>
<td>27</td>
<td>26</td>
<td>22</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>.07</td>
<td>48</td>
<td>45</td>
<td>44</td>
<td>19</td>
<td>41</td>
<td>24</td>
<td>40</td>
<td>37</td>
<td>32</td>
<td>26</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>.10</td>
<td>52</td>
<td>50</td>
<td>49</td>
<td>26</td>
<td>51</td>
<td>36</td>
<td>50</td>
<td>49</td>
<td>42</td>
<td>33</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td><strong>33% Spuriously High Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>23</td>
<td>24</td>
<td>17</td>
<td>02</td>
<td>10</td>
<td>00</td>
<td>04</td>
<td>00</td>
<td>06</td>
<td>12</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>40</td>
<td>33</td>
<td>27</td>
<td>07</td>
<td>25</td>
<td>00</td>
<td>15</td>
<td>00</td>
<td>28</td>
<td>27</td>
<td>00</td>
<td>04</td>
</tr>
<tr>
<td>.01</td>
<td>45</td>
<td>45</td>
<td>43</td>
<td>12</td>
<td>30</td>
<td>01</td>
<td>27</td>
<td>06</td>
<td>37</td>
<td>34</td>
<td>00</td>
<td>09</td>
</tr>
<tr>
<td>.02</td>
<td>61</td>
<td>54</td>
<td>52</td>
<td>17</td>
<td>44</td>
<td>05</td>
<td>41</td>
<td>17</td>
<td>50</td>
<td>46</td>
<td>01</td>
<td>17</td>
</tr>
<tr>
<td>.03</td>
<td>67</td>
<td>59</td>
<td>58</td>
<td>22</td>
<td>50</td>
<td>16</td>
<td>47</td>
<td>24</td>
<td>60</td>
<td>52</td>
<td>02</td>
<td>24</td>
</tr>
<tr>
<td>.04</td>
<td>71</td>
<td>64</td>
<td>63</td>
<td>25</td>
<td>56</td>
<td>23</td>
<td>55</td>
<td>32</td>
<td>65</td>
<td>57</td>
<td>03</td>
<td>37</td>
</tr>
<tr>
<td>.05</td>
<td>72</td>
<td>67</td>
<td>66</td>
<td>31</td>
<td>62</td>
<td>30</td>
<td>59</td>
<td>37</td>
<td>69</td>
<td>61</td>
<td>03</td>
<td>37</td>
</tr>
<tr>
<td>.07</td>
<td>77</td>
<td>71</td>
<td>70</td>
<td>37</td>
<td>66</td>
<td>42</td>
<td>68</td>
<td>47</td>
<td>74</td>
<td>67</td>
<td>07</td>
<td>47</td>
</tr>
<tr>
<td>.10</td>
<td>81</td>
<td>75</td>
<td>75</td>
<td>46</td>
<td>75</td>
<td>57</td>
<td>76</td>
<td>60</td>
<td>81</td>
<td>73</td>
<td>19</td>
<td>57</td>
</tr>
</tbody>
</table>
Table 2
Selected ROC Points for Spuriously High Response Patterns Generated from the 10-30% Ability Range

<table>
<thead>
<tr>
<th>False Alarm Rate (%)</th>
<th>LR, LR, LR, z_p, z, F1, F2, S, T2, T4, JK, O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>17% Spuriously High Treatment</td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>02 01 00 00 00 00 00 00 00 00 01 00 00 00</td>
</tr>
<tr>
<td>.005</td>
<td>09 07 07 01 05 00 03 00 05 04 00 01 00 01</td>
</tr>
<tr>
<td>.01</td>
<td>14 14 14 04 09 00 06 00 07 07 00 03 00 03</td>
</tr>
<tr>
<td>.02</td>
<td>26 25 22 06 14 01 11 04 14 12 01 05 05 05</td>
</tr>
<tr>
<td>.03</td>
<td>31 29 29 08 19 03 14 06 20 16 02 07 05 08</td>
</tr>
<tr>
<td>.04</td>
<td>34 33 33 10 23 06 18 10 24 20 02 10 06 10</td>
</tr>
<tr>
<td>.05</td>
<td>40 36 37 12 27 09 21 12 27 23 03 14 13 13</td>
</tr>
<tr>
<td>.07</td>
<td>46 43 43 16 34 14 27 18 33 28 06 20 11 20</td>
</tr>
<tr>
<td>.10</td>
<td>52 50 51 23 43 24 37 28 42 34 14 27 14 27</td>
</tr>
<tr>
<td>33% Spuriously High Treatment</td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>16 16 13 00 04 00 00 00 03 09 00 01 00 01</td>
</tr>
<tr>
<td>.005</td>
<td>31 27 23 03 14 00 07 00 20 23 00 06 06 06</td>
</tr>
<tr>
<td>.01</td>
<td>37 40 39 05 20 00 15 01 28 29 00 10 10 10</td>
</tr>
<tr>
<td>.02</td>
<td>53 50 50 08 30 03 27 06 42 41 00 20 00 20</td>
</tr>
<tr>
<td>.03</td>
<td>61 56 57 12 37 08 34 10 51 47 00 27 00 20</td>
</tr>
<tr>
<td>.04</td>
<td>65 63 62 14 42 12 42 16 58 53 00 34 34 34</td>
</tr>
<tr>
<td>.05</td>
<td>68 66 65 19 49 17 46 20 62 58 00 40 00 30</td>
</tr>
<tr>
<td>.07</td>
<td>73 70 70 25 54 28 56 29 67 63 05 05 05 05</td>
</tr>
<tr>
<td>.10</td>
<td>78 74 75 33 64 44 67 41 74 70 18 60 18 60</td>
</tr>
</tbody>
</table>
Table 3
Selected ROC Points for Spuriously High Response Patterns Generated from the 31-48% Ability Range

<table>
<thead>
<tr>
<th>Rate</th>
<th>LRₚ</th>
<th>LRₚ⁺</th>
<th>LRₚ⁻</th>
<th>zᵢₚ</th>
<th>zᵢ⁻</th>
<th>F₁</th>
<th>F₂</th>
<th>S</th>
<th>T₂</th>
<th>T₄</th>
<th>JK</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>03</td>
<td>04</td>
<td>03</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>04</td>
<td>04</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>06</td>
<td>07</td>
<td>08</td>
<td>02</td>
<td>06</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>06</td>
<td>06</td>
<td>00</td>
<td>01</td>
</tr>
<tr>
<td>.02</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>03</td>
<td>09</td>
<td>05</td>
<td>00</td>
<td>00</td>
<td>12</td>
<td>12</td>
<td>00</td>
<td>05</td>
</tr>
<tr>
<td>.03</td>
<td>20</td>
<td>19</td>
<td>18</td>
<td>05</td>
<td>14</td>
<td>03</td>
<td>07</td>
<td>02</td>
<td>17</td>
<td>15</td>
<td>00</td>
<td>08</td>
</tr>
<tr>
<td>.04</td>
<td>24</td>
<td>23</td>
<td>24</td>
<td>06</td>
<td>17</td>
<td>06</td>
<td>10</td>
<td>03</td>
<td>21</td>
<td>18</td>
<td>00</td>
<td>10</td>
</tr>
<tr>
<td>.05</td>
<td>29</td>
<td>28</td>
<td>28</td>
<td>07</td>
<td>20</td>
<td>07</td>
<td>13</td>
<td>04</td>
<td>23</td>
<td>22</td>
<td>00</td>
<td>13</td>
</tr>
<tr>
<td>.07</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td>10</td>
<td>25</td>
<td>12</td>
<td>12</td>
<td>07</td>
<td>29</td>
<td>26</td>
<td>01</td>
<td>20</td>
</tr>
<tr>
<td>.10</td>
<td>43</td>
<td>42</td>
<td>43</td>
<td>15</td>
<td>33</td>
<td>18</td>
<td>26</td>
<td>12</td>
<td>36</td>
<td>32</td>
<td>07</td>
<td>29</td>
</tr>
</tbody>
</table>

33% Spuriously High Treatment

<table>
<thead>
<tr>
<th>Rate</th>
<th>LRₚ</th>
<th>LRₚ⁺</th>
<th>LRₚ⁻</th>
<th>zᵢₚ</th>
<th>zᵢ⁻</th>
<th>F₁</th>
<th>F₂</th>
<th>S</th>
<th>T₂</th>
<th>T₄</th>
<th>JK</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>06</td>
<td>07</td>
<td>07</td>
<td>00</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>02</td>
<td>06</td>
<td>00</td>
<td>01</td>
</tr>
<tr>
<td>.005</td>
<td>17</td>
<td>16</td>
<td>14</td>
<td>01</td>
<td>07</td>
<td>00</td>
<td>03</td>
<td>00</td>
<td>12</td>
<td>16</td>
<td>00</td>
<td>05</td>
</tr>
<tr>
<td>.01</td>
<td>22</td>
<td>27</td>
<td>26</td>
<td>02</td>
<td>10</td>
<td>00</td>
<td>08</td>
<td>00</td>
<td>18</td>
<td>22</td>
<td>00</td>
<td>08</td>
</tr>
<tr>
<td>.02</td>
<td>39</td>
<td>36</td>
<td>37</td>
<td>04</td>
<td>17</td>
<td>04</td>
<td>16</td>
<td>02</td>
<td>27</td>
<td>32</td>
<td>00</td>
<td>17</td>
</tr>
<tr>
<td>.03</td>
<td>48</td>
<td>43</td>
<td>45</td>
<td>05</td>
<td>22</td>
<td>08</td>
<td>21</td>
<td>05</td>
<td>36</td>
<td>38</td>
<td>00</td>
<td>23</td>
</tr>
<tr>
<td>.04</td>
<td>53</td>
<td>51</td>
<td>49</td>
<td>07</td>
<td>27</td>
<td>12</td>
<td>27</td>
<td>07</td>
<td>41</td>
<td>43</td>
<td>00</td>
<td>29</td>
</tr>
<tr>
<td>.05</td>
<td>56</td>
<td>55</td>
<td>54</td>
<td>09</td>
<td>33</td>
<td>16</td>
<td>31</td>
<td>09</td>
<td>45</td>
<td>47</td>
<td>00</td>
<td>34</td>
</tr>
<tr>
<td>.07</td>
<td>61</td>
<td>61</td>
<td>61</td>
<td>13</td>
<td>37</td>
<td>23</td>
<td>40</td>
<td>14</td>
<td>50</td>
<td>53</td>
<td>07</td>
<td>44</td>
</tr>
<tr>
<td>.10</td>
<td>71</td>
<td>67</td>
<td>68</td>
<td>20</td>
<td>46</td>
<td>36</td>
<td>51</td>
<td>22</td>
<td>59</td>
<td>60</td>
<td>19</td>
<td>53</td>
</tr>
</tbody>
</table>
Table 4
Selected ROC Points for Spuriously High Response Patterns Generated from the 49-64% Ability Range

<table>
<thead>
<tr>
<th>False Alarm Rate</th>
<th>LRₚ</th>
<th>LR₁</th>
<th>z₀</th>
<th>z₁</th>
<th>F₁</th>
<th>F₂</th>
<th>S</th>
<th>T₂</th>
<th>T₄</th>
<th>JK</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>17% Spuriously High Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>02</td>
<td>03</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>02</td>
<td>01</td>
<td>01</td>
<td>00</td>
<td>03</td>
<td>01</td>
<td>01</td>
<td>00</td>
<td>04</td>
<td>04</td>
<td>00</td>
</tr>
<tr>
<td>.02</td>
<td>07</td>
<td>06</td>
<td>03</td>
<td>01</td>
<td>05</td>
<td>01</td>
<td>03</td>
<td>00</td>
<td>07</td>
<td>08</td>
<td>00</td>
</tr>
<tr>
<td>.03</td>
<td>11</td>
<td>09</td>
<td>05</td>
<td>01</td>
<td>08</td>
<td>04</td>
<td>04</td>
<td>00</td>
<td>11</td>
<td>11</td>
<td>00</td>
</tr>
<tr>
<td>.04</td>
<td>14</td>
<td>13</td>
<td>07</td>
<td>02</td>
<td>10</td>
<td>06</td>
<td>07</td>
<td>01</td>
<td>14</td>
<td>14</td>
<td>00</td>
</tr>
<tr>
<td>.05</td>
<td>18</td>
<td>16</td>
<td>09</td>
<td>03</td>
<td>13</td>
<td>08</td>
<td>08</td>
<td>01</td>
<td>16</td>
<td>17</td>
<td>00</td>
</tr>
<tr>
<td>.07</td>
<td>25</td>
<td>23</td>
<td>14</td>
<td>06</td>
<td>17</td>
<td>11</td>
<td>13</td>
<td>03</td>
<td>20</td>
<td>21</td>
<td>01</td>
</tr>
<tr>
<td>.10</td>
<td>33</td>
<td>30</td>
<td>23</td>
<td>09</td>
<td>23</td>
<td>16</td>
<td>19</td>
<td>05</td>
<td>26</td>
<td>27</td>
<td>07</td>
</tr>
<tr>
<td>33% Spuriously High Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>01</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>02</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>05</td>
<td>04</td>
<td>02</td>
<td>03</td>
<td>03</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>05</td>
<td>07</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>08</td>
<td>10</td>
<td>05</td>
<td>00</td>
<td>04</td>
<td>02</td>
<td>04</td>
<td>00</td>
<td>07</td>
<td>10</td>
<td>00</td>
</tr>
<tr>
<td>.02</td>
<td>19</td>
<td>16</td>
<td>09</td>
<td>01</td>
<td>07</td>
<td>07</td>
<td>08</td>
<td>01</td>
<td>12</td>
<td>17</td>
<td>00</td>
</tr>
<tr>
<td>.03</td>
<td>28</td>
<td>23</td>
<td>14</td>
<td>03</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>02</td>
<td>16</td>
<td>20</td>
<td>00</td>
</tr>
<tr>
<td>.04</td>
<td>34</td>
<td>32</td>
<td>18</td>
<td>03</td>
<td>12</td>
<td>14</td>
<td>15</td>
<td>03</td>
<td>20</td>
<td>25</td>
<td>00</td>
</tr>
<tr>
<td>.05</td>
<td>37</td>
<td>37</td>
<td>21</td>
<td>05</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>04</td>
<td>23</td>
<td>29</td>
<td>00</td>
</tr>
<tr>
<td>.07</td>
<td>48</td>
<td>45</td>
<td>29</td>
<td>08</td>
<td>19</td>
<td>23</td>
<td>23</td>
<td>07</td>
<td>28</td>
<td>35</td>
<td>03</td>
</tr>
<tr>
<td>.10</td>
<td>60</td>
<td>55</td>
<td>41</td>
<td>13</td>
<td>25</td>
<td>31</td>
<td>31</td>
<td>12</td>
<td>35</td>
<td>40</td>
<td>11</td>
</tr>
</tbody>
</table>
### Table 5
Selected ROC Points for Spuriously Low Response Patterns Generated from the 31-48% Ability Range

<table>
<thead>
<tr>
<th>False Alarm Rate</th>
<th>LRₚ</th>
<th>LR₁</th>
<th>LR₂</th>
<th>z₁</th>
<th>F₁</th>
<th>F₂</th>
<th>S</th>
<th>T₂</th>
<th>T₄</th>
<th>JK</th>
<th>O/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>17% Spuriously Low Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>05</td>
<td>01</td>
<td>01</td>
<td>03</td>
<td>02</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>02</td>
<td>02</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>09</td>
<td>03</td>
<td>03</td>
<td>05</td>
<td>04</td>
<td>01</td>
<td>02</td>
<td>00</td>
<td>03</td>
<td>03</td>
<td>01</td>
</tr>
<tr>
<td>.02</td>
<td>15</td>
<td>06</td>
<td>07</td>
<td>08</td>
<td>07</td>
<td>02</td>
<td>04</td>
<td>00</td>
<td>06</td>
<td>07</td>
<td>01</td>
</tr>
<tr>
<td>.03</td>
<td>18</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>04</td>
<td>05</td>
<td>01</td>
<td>09</td>
<td>09</td>
<td>02</td>
</tr>
<tr>
<td>.04</td>
<td>21</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>07</td>
<td>07</td>
<td>03</td>
<td>12</td>
<td>12</td>
<td>03</td>
</tr>
<tr>
<td>.05</td>
<td>24</td>
<td>17</td>
<td>18</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>09</td>
<td>04</td>
<td>14</td>
<td>14</td>
<td>05</td>
</tr>
<tr>
<td>.07</td>
<td>29</td>
<td>22</td>
<td>23</td>
<td>21</td>
<td>19</td>
<td>17</td>
<td>12</td>
<td>07</td>
<td>18</td>
<td>17</td>
<td>07</td>
</tr>
<tr>
<td>.10</td>
<td>35</td>
<td>28</td>
<td>28</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>17</td>
<td>11</td>
<td>23</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>33% Spuriously Low Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>07</td>
<td>01</td>
<td>01</td>
<td>01</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>14</td>
<td>03</td>
<td>04</td>
<td>07</td>
<td>05</td>
<td>00</td>
<td>04</td>
<td>00</td>
<td>03</td>
<td>04</td>
<td>01</td>
</tr>
<tr>
<td>.01</td>
<td>22</td>
<td>08</td>
<td>09</td>
<td>12</td>
<td>10</td>
<td>02</td>
<td>07</td>
<td>00</td>
<td>05</td>
<td>07</td>
<td>02</td>
</tr>
<tr>
<td>.02</td>
<td>30</td>
<td>14</td>
<td>16</td>
<td>18</td>
<td>15</td>
<td>05</td>
<td>11</td>
<td>03</td>
<td>09</td>
<td>11</td>
<td>04</td>
</tr>
<tr>
<td>.03</td>
<td>36</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>20</td>
<td>09</td>
<td>13</td>
<td>06</td>
<td>14</td>
<td>15</td>
<td>07</td>
</tr>
<tr>
<td>.04</td>
<td>41</td>
<td>24</td>
<td>26</td>
<td>27</td>
<td>23</td>
<td>13</td>
<td>17</td>
<td>10</td>
<td>16</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>.05</td>
<td>45</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>26</td>
<td>17</td>
<td>19</td>
<td>11</td>
<td>19</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>.07</td>
<td>51</td>
<td>36</td>
<td>37</td>
<td>36</td>
<td>32</td>
<td>27</td>
<td>24</td>
<td>17</td>
<td>22</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>.10</td>
<td>59</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>38</td>
<td>36</td>
<td>31</td>
<td>25</td>
<td>29</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>False Alarm Rate</td>
<td>LRₚ</td>
<td>LR₁</td>
<td>LR₂</td>
<td>z₀</td>
<td>z₀</td>
<td>F₁</td>
<td>F₂</td>
<td>S</td>
<td>T₂</td>
<td>T₄</td>
<td>JK</td>
</tr>
<tr>
<td>------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td><strong>17% Spuriously Low Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>13</td>
<td>02</td>
<td>02</td>
<td>00</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>22</td>
<td>09</td>
<td>08</td>
<td>04</td>
<td>04</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>05</td>
<td>05</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>26</td>
<td>14</td>
<td>13</td>
<td>07</td>
<td>09</td>
<td>03</td>
<td>03</td>
<td>00</td>
<td>08</td>
<td>07</td>
<td>00</td>
</tr>
<tr>
<td>.02</td>
<td>32</td>
<td>21</td>
<td>20</td>
<td>11</td>
<td>14</td>
<td>09</td>
<td>07</td>
<td>01</td>
<td>13</td>
<td>13</td>
<td>00</td>
</tr>
<tr>
<td>.03</td>
<td>34</td>
<td>26</td>
<td>25</td>
<td>16</td>
<td>19</td>
<td>17</td>
<td>10</td>
<td>03</td>
<td>19</td>
<td>16</td>
<td>00</td>
</tr>
<tr>
<td>.04</td>
<td>38</td>
<td>30</td>
<td>28</td>
<td>19</td>
<td>22</td>
<td>21</td>
<td>13</td>
<td>04</td>
<td>22</td>
<td>20</td>
<td>00</td>
</tr>
<tr>
<td>.05</td>
<td>41</td>
<td>33</td>
<td>31</td>
<td>23</td>
<td>25</td>
<td>24</td>
<td>15</td>
<td>05</td>
<td>26</td>
<td>22</td>
<td>00</td>
</tr>
<tr>
<td>.07</td>
<td>46</td>
<td>37</td>
<td>35</td>
<td>29</td>
<td>31</td>
<td>31</td>
<td>20</td>
<td>08</td>
<td>29</td>
<td>27</td>
<td>03</td>
</tr>
<tr>
<td>.10</td>
<td>51</td>
<td>42</td>
<td>40</td>
<td>37</td>
<td>38</td>
<td>34</td>
<td>28</td>
<td>13</td>
<td>36</td>
<td>32</td>
<td>09</td>
</tr>
<tr>
<td><strong>33% Spuriously Low Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>24</td>
<td>09</td>
<td>08</td>
<td>02</td>
<td>08</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>03</td>
<td>05</td>
<td>00</td>
</tr>
<tr>
<td>.005</td>
<td>34</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>00</td>
<td>06</td>
<td>00</td>
<td>14</td>
<td>12</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>43</td>
<td>25</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>22</td>
<td>11</td>
<td>01</td>
<td>18</td>
<td>17</td>
<td>00</td>
</tr>
<tr>
<td>.02</td>
<td>51</td>
<td>33</td>
<td>32</td>
<td>31</td>
<td>29</td>
<td>12</td>
<td>17</td>
<td>04</td>
<td>26</td>
<td>25</td>
<td>00</td>
</tr>
<tr>
<td>.03</td>
<td>55</td>
<td>39</td>
<td>38</td>
<td>37</td>
<td>36</td>
<td>22</td>
<td>22</td>
<td>07</td>
<td>33</td>
<td>29</td>
<td>01</td>
</tr>
<tr>
<td>.04</td>
<td>58</td>
<td>43</td>
<td>41</td>
<td>41</td>
<td>39</td>
<td>29</td>
<td>26</td>
<td>10</td>
<td>38</td>
<td>34</td>
<td>01</td>
</tr>
<tr>
<td>.05</td>
<td>61</td>
<td>46</td>
<td>45</td>
<td>45</td>
<td>43</td>
<td>35</td>
<td>29</td>
<td>13</td>
<td>41</td>
<td>38</td>
<td>02</td>
</tr>
<tr>
<td>.07</td>
<td>66</td>
<td>51</td>
<td>50</td>
<td>52</td>
<td>49</td>
<td>44</td>
<td>36</td>
<td>19</td>
<td>45</td>
<td>44</td>
<td>06</td>
</tr>
<tr>
<td>.10</td>
<td>71</td>
<td>58</td>
<td>56</td>
<td>60</td>
<td>57</td>
<td>52</td>
<td>45</td>
<td>27</td>
<td>53</td>
<td>51</td>
<td>14</td>
</tr>
</tbody>
</table>
Table 7
Selected ROC Points for Spuriously Low Response Patterns Generated from the 65-92% Ability Range

<table>
<thead>
<tr>
<th>False Alarm Rate</th>
<th>( LR_p )</th>
<th>( LR_z )</th>
<th>( LR_z' )</th>
<th>( z_p )</th>
<th>( z )</th>
<th>( F1 )</th>
<th>( F2 )</th>
<th>( S )</th>
<th>( T2 )</th>
<th>( T4 )</th>
<th>( JK )</th>
<th>( J/E )</th>
</tr>
</thead>
<tbody>
<tr>
<td>17% Spuriously Low Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>35</td>
<td>15</td>
<td>14</td>
<td>01</td>
<td>05</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>01</td>
<td>03</td>
<td>00</td>
<td>02</td>
</tr>
<tr>
<td>.005</td>
<td>45</td>
<td>33</td>
<td>31</td>
<td>09</td>
<td>11</td>
<td>11</td>
<td>05</td>
<td>01</td>
<td>11</td>
<td>09</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.01</td>
<td>51</td>
<td>41</td>
<td>38</td>
<td>15</td>
<td>19</td>
<td>24</td>
<td>11</td>
<td>05</td>
<td>15</td>
<td>14</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.02</td>
<td>57</td>
<td>46</td>
<td>45</td>
<td>20</td>
<td>27</td>
<td>40</td>
<td>20</td>
<td>13</td>
<td>25</td>
<td>21</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.03</td>
<td>60</td>
<td>50</td>
<td>49</td>
<td>25</td>
<td>34</td>
<td>45</td>
<td>24</td>
<td>13</td>
<td>32</td>
<td>26</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.04</td>
<td>63</td>
<td>53</td>
<td>51</td>
<td>31</td>
<td>38</td>
<td>53</td>
<td>30</td>
<td>24</td>
<td>36</td>
<td>30</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.05</td>
<td>65</td>
<td>55</td>
<td>53</td>
<td>35</td>
<td>42</td>
<td>57</td>
<td>34</td>
<td>27</td>
<td>39</td>
<td>35</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>.07</td>
<td>68</td>
<td>59</td>
<td>57</td>
<td>42</td>
<td>50</td>
<td>61</td>
<td>41</td>
<td>33</td>
<td>43</td>
<td>40</td>
<td>04</td>
<td>10</td>
</tr>
<tr>
<td>.10</td>
<td>71</td>
<td>63</td>
<td>61</td>
<td>51</td>
<td>51</td>
<td>65</td>
<td>50</td>
<td>41</td>
<td>53</td>
<td>48</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>33% Spuriously Low Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>53</td>
<td>31</td>
<td>29</td>
<td>04</td>
<td>26</td>
<td>00</td>
<td>03</td>
<td>00</td>
<td>15</td>
<td>22</td>
<td>00</td>
<td>02</td>
</tr>
<tr>
<td>.005</td>
<td>61</td>
<td>44</td>
<td>41</td>
<td>25</td>
<td>39</td>
<td>05</td>
<td>21</td>
<td>02</td>
<td>40</td>
<td>39</td>
<td>00</td>
<td>09</td>
</tr>
<tr>
<td>.01</td>
<td>67</td>
<td>53</td>
<td>51</td>
<td>34</td>
<td>52</td>
<td>20</td>
<td>34</td>
<td>09</td>
<td>47</td>
<td>46</td>
<td>00</td>
<td>15</td>
</tr>
<tr>
<td>.02</td>
<td>72</td>
<td>59</td>
<td>57</td>
<td>45</td>
<td>61</td>
<td>42</td>
<td>47</td>
<td>21</td>
<td>59</td>
<td>57</td>
<td>00</td>
<td>25</td>
</tr>
<tr>
<td>.03</td>
<td>75</td>
<td>63</td>
<td>61</td>
<td>52</td>
<td>67</td>
<td>54</td>
<td>54</td>
<td>29</td>
<td>67</td>
<td>67</td>
<td>00</td>
<td>30</td>
</tr>
<tr>
<td>.04</td>
<td>77</td>
<td>66</td>
<td>63</td>
<td>57</td>
<td>70</td>
<td>60</td>
<td>61</td>
<td>36</td>
<td>71</td>
<td>67</td>
<td>00</td>
<td>38</td>
</tr>
<tr>
<td>.05</td>
<td>78</td>
<td>68</td>
<td>66</td>
<td>61</td>
<td>74</td>
<td>64</td>
<td>64</td>
<td>40</td>
<td>74</td>
<td>70</td>
<td>00</td>
<td>43</td>
</tr>
<tr>
<td>.07</td>
<td>81</td>
<td>72</td>
<td>70</td>
<td>69</td>
<td>79</td>
<td>71</td>
<td>71</td>
<td>49</td>
<td>79</td>
<td>75</td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td>.10</td>
<td>84</td>
<td>75</td>
<td>74</td>
<td>77</td>
<td>84</td>
<td>79</td>
<td>79</td>
<td>58</td>
<td>84</td>
<td>80</td>
<td>33</td>
<td>61</td>
</tr>
</tbody>
</table>
Table 4
Selected Y% Points for Spuriously Low
Response Patterns Generated from the 45-100% Ability Range

<table>
<thead>
<tr>
<th>False Alarm</th>
<th>( LR_0 )</th>
<th>( LR_1 )</th>
<th>( LR_2 )</th>
<th>( z )</th>
<th>( F_0 )</th>
<th>( F_1 )</th>
<th>( T_0 )</th>
<th>( T_1 )</th>
<th>( J_0 )</th>
<th>( J_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>45</td>
<td>22</td>
<td>22</td>
<td>04</td>
<td>04</td>
<td>11</td>
<td>00</td>
<td>00</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>.005</td>
<td>55</td>
<td>45</td>
<td>45</td>
<td>13</td>
<td>13</td>
<td>17</td>
<td>09</td>
<td>09</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>.01</td>
<td>60</td>
<td>49</td>
<td>49</td>
<td>18</td>
<td>18</td>
<td>20</td>
<td>18</td>
<td>18</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>.02</td>
<td>57</td>
<td>54</td>
<td>54</td>
<td>26</td>
<td>26</td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>.03</td>
<td>59</td>
<td>58</td>
<td>58</td>
<td>32</td>
<td>32</td>
<td>37</td>
<td>40</td>
<td>40</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>.04</td>
<td>61</td>
<td>60</td>
<td>60</td>
<td>37</td>
<td>37</td>
<td>41</td>
<td>41</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>.05</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>40</td>
<td>40</td>
<td>46</td>
<td>40</td>
<td>40</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>.07</td>
<td>74</td>
<td>65</td>
<td>65</td>
<td>48</td>
<td>48</td>
<td>71</td>
<td>51</td>
<td>51</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>.10</td>
<td>77</td>
<td>68</td>
<td>68</td>
<td>58</td>
<td>58</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>66</td>
<td>66</td>
</tr>
</tbody>
</table>

17% Spuriously Low Treatment

<table>
<thead>
<tr>
<th>False Alarm</th>
<th>( LR_0 )</th>
<th>( LR_1 )</th>
<th>( LR_2 )</th>
<th>( z )</th>
<th>( F_0 )</th>
<th>( F_1 )</th>
<th>( T_0 )</th>
<th>( T_1 )</th>
<th>( J_0 )</th>
<th>( J_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>54</td>
<td>42</td>
<td>42</td>
<td>24</td>
<td>24</td>
<td>32</td>
<td>02</td>
<td>02</td>
<td>02</td>
<td>02</td>
</tr>
<tr>
<td>.005</td>
<td>72</td>
<td>55</td>
<td>55</td>
<td>27</td>
<td>27</td>
<td>49</td>
<td>17</td>
<td>17</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>.01</td>
<td>76</td>
<td>61</td>
<td>61</td>
<td>39</td>
<td>39</td>
<td>62</td>
<td>36</td>
<td>36</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>.02</td>
<td>81</td>
<td>67</td>
<td>67</td>
<td>51</td>
<td>51</td>
<td>71</td>
<td>59</td>
<td>59</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>.03</td>
<td>83</td>
<td>70</td>
<td>70</td>
<td>59</td>
<td>59</td>
<td>77</td>
<td>69</td>
<td>69</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>.04</td>
<td>85</td>
<td>72</td>
<td>72</td>
<td>64</td>
<td>64</td>
<td>79</td>
<td>74</td>
<td>74</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>.05</td>
<td>86</td>
<td>74</td>
<td>74</td>
<td>68</td>
<td>68</td>
<td>82</td>
<td>78</td>
<td>78</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>.07</td>
<td>87</td>
<td>77</td>
<td>77</td>
<td>75</td>
<td>75</td>
<td>86</td>
<td>84</td>
<td>84</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>.10</td>
<td>90</td>
<td>79</td>
<td>79</td>
<td>82</td>
<td>82</td>
<td>90</td>
<td>87</td>
<td>87</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

33% Spuriously Low Treatment

<table>
<thead>
<tr>
<th>False Alarm</th>
<th>( LR_0 )</th>
<th>( LR_1 )</th>
<th>( LR_2 )</th>
<th>( z )</th>
<th>( F_0 )</th>
<th>( F_1 )</th>
<th>( T_0 )</th>
<th>( T_1 )</th>
<th>( J_0 )</th>
<th>( J_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>64</td>
<td>42</td>
<td>42</td>
<td>24</td>
<td>24</td>
<td>32</td>
<td>02</td>
<td>02</td>
<td>02</td>
<td>02</td>
</tr>
<tr>
<td>.005</td>
<td>72</td>
<td>55</td>
<td>55</td>
<td>27</td>
<td>27</td>
<td>49</td>
<td>17</td>
<td>17</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>.01</td>
<td>76</td>
<td>61</td>
<td>61</td>
<td>39</td>
<td>39</td>
<td>62</td>
<td>36</td>
<td>36</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>.02</td>
<td>81</td>
<td>67</td>
<td>67</td>
<td>51</td>
<td>51</td>
<td>71</td>
<td>59</td>
<td>59</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>.03</td>
<td>83</td>
<td>70</td>
<td>70</td>
<td>59</td>
<td>59</td>
<td>77</td>
<td>69</td>
<td>69</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>.04</td>
<td>85</td>
<td>72</td>
<td>72</td>
<td>64</td>
<td>64</td>
<td>79</td>
<td>74</td>
<td>74</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>.05</td>
<td>86</td>
<td>74</td>
<td>74</td>
<td>68</td>
<td>68</td>
<td>82</td>
<td>78</td>
<td>78</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>.07</td>
<td>87</td>
<td>77</td>
<td>77</td>
<td>75</td>
<td>75</td>
<td>86</td>
<td>84</td>
<td>84</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>.10</td>
<td>90</td>
<td>79</td>
<td>79</td>
<td>82</td>
<td>82</td>
<td>90</td>
<td>87</td>
<td>87</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>
Figure captions


2. Information functions for dichotomous and polytomous scorings of the Arithmetic Reasoning test.
Appendix 1

Goodness of Fit of Arithmetic Reasoning COCCs Estimated from a Sample of
N=2,978 and Evaluated Using the Entire Sample of N=11,914.
ITEM 3

PROBABILITY

THETA

-3.0 0 3.0

THETA

-3.0 0 3.0
ITEM 4

![Graphs showing probability distribution](image)
ITEM 7
ITEM 22

The diagrams show plots of some data with respect to theta. The vertical axes represent different values, possibly probability or another measurable quantity, while the horizontal axes are labeled 'THETA'. Each plot contains a set of data points with error bars and a fitted curve. The plots are arranged in a 2x2 grid.
ITEM 23
ITEM 25

The diagrams show the relationship between the variable $\Theta$ and another variable, with data points and trend lines illustrating the distribution and trends over the range of $\Theta$ from $-3.0$ to $3.0$. The plots depict the variation and potential anomalies or patterns in the data.
ITEM 30

The diagrams show the probability distributions for different values of THETA.

- The top left diagram illustrates a decreasing trend with THETA.
- The top right diagram shows a complex distribution with a peak at around THETA = 0.
- The bottom left diagram also displays a peak but with a more pronounced curve.
- The bottom right diagram exhibits a similar pattern but with slightly different values.

The y-axis represents the probability, ranging from 0.0 to 1.0, while the x-axis represents THETA, ranging from -3.0 to 3.0.
Appendix 2

Estimated COCCs, Simulation COCCs, and Empirical Proportions from Estimation Sample.
ITEM 1

PROBABILITY

 THETA

THETA

PROBABILITY

0.0

0.5

1.0
ITEM 3

![Graphs showing probability distributions for different values of THETA.](image-url)
ITEM 4

The diagrams depict the probability distribution of a variable $\theta$ over the range $-3.0$ to $3.0$. The distributions are shown in four quadrants, each with varying patterns of probability density and confidence intervals.
ITEM 14

![Graphs showing probability vs. theta with error bars and curves.](image)
ITEM 16

PROBABILITY

THETA

-3.0 0 3.0

-3.0 0 3.0
ITEM 19
ITEM 23

PROBABILITY

-3.0 0 3.0

THETA

-3.0 0 3.0
ITEM 24

![Graphs showing probability distribution over theta values.](image-url)
ITEM 29

PROBABILITY

THETA

PROBABILITY

THETA

PROBABILITY

THETA

PROBABILITY

THETA
Dr. Norman Cliff  
Department of Psychology  
Univ. of So. California  
University Park  
Los Angeles, CA 90007

Director,  
Manpower Support and  
Readiness Program  
Center for Naval Analysis  
2000 North Beauregard Street  
Alexandria, VA 22311

Dr. Stanley Collyer  
Office of Naval Technology  
Code 222  
800 N. Quincy Street  
Arlington, VA 22217-5000

Dr. Hans Crombag  
University of Leyden  
Education Research Center  
Boerhaavelaan 2  
2334 EN Leyden  
The NETHERLANDS

Mr. Timothy Davey  
University of Illinois  
Educational Psychology  
Urbana, IL 61801

Dr. Dattprasad Divgi  
Center for Naval Analysis  
4401 Ford Avenue  
P.O. Box 16268  
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong  
Bell Communications Research  
6 Corporate Place  
PYA-1k226  
Piscataway, NJ 08854

Dr. Fritz Drasgow  
University of Illinois  
Department of Psychology  
603 E. Daniel St.  
Champaign, IL 61820

Defense Technical  
Information Center  
Cameron Station, Bldg 5  
Alexandria, VA 22314  
Attn: TC  
(12 Copies)

Dr. Stephen Dunbar  
Lindquist Center  
for Measurement  
University of Iowa  
Iowa City, IA 52242

Dr. James A. Earles  
Air Force Human Resources Lab  
Brooks AFB, TX 78235

Dr. Kent Eaton  
Army Research Institute  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Dr. John M. Eddins  
University of Illinois  
252 Engineering Research Laboratory  
103 South Mathews Street  
Urbana, IL 61801

Dr. Susan Embretson  
University of Kansas  
Psychology Department  
426 Fraser  
Lawrence, KS 66045

Dr. George Englehard, Jr.  
Division of Educational Studies  
Emory University  
201 Fishburne Bldg.  
Atlanta, GA 30322

Dr. Benjamin A. Fairbank  
Performance Metrics, Inc.  
5825 Callaghan  
Suite 225  
San Antonio, TX 78228

Dr. Pat Federico  
Code 511  
NPRDC  
San Diego, CA 92152-6800
Dr. James Lumsden  
Department of Psychology  
University of Western Australia  
 Nedlands W.A. 6009  
 AUSTRALIA

Dr. Milton Haier  
Center for Naval Analysis  
4401 Ford Avenue  
P.O. Box 16268  
Alexandria, VA 22302-0268

Dr. William L. Maloy  
Chief of Naval Education and Training  
Naval Air Station  
Pensacola, FL 32508

Dr. Gary Marco  
Stop 31-E  
Educational Testing Service  
Princeton, NJ 08451

Dr. Clessen Martin  
Army Research Institute  
5001 Eisenhower Blvd.  
Alexandria, VA 22333

Dr. James McBride  
Psychological Corporation  
c/o Harcourt, Brace,  
Jovanovich Inc.  
1250 West 6th Street  
San Diego, CA 92101

Dr. Clarence McCormick  
HQ, MEPCOM  
MEPCT-P  
2500 Green Bay Road  
North Chicago, IL 60064

Dr. Robert McKinley  
Educational Testing Service  
20-P  
Princeton, NJ 08541

Dr. James McMichael  
Technical Director  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. Barbara Means  
Human Resources Research Organization  
1100 South Washington  
Alexandria, VA 22314

Dr. Robert Mislevy  
Educational Testing Service  
Princeton, NJ 08541

Dr. William Montague  
NPRDC Code 13  
San Diego, CA 92152-6800

Ms. Kathleen Moreno  
Navy Personnel R&D Center  
Code 62  
San Diego, CA 92152-6800

Headquarters, Marine Corps  
Code MPI-20  
Washington, DC 20380

Dr. W. Alan Nicewander  
University of Oklahoma  
Department of Psychology  
Oklahoma City, OK 73069

Deputy Technical Director  
NPRDC Code 01A  
San Diego, CA 92152-6800

Director, Training Laboratory,  
NPRDC (Code 05)  
San Diego, CA 92152-6800

Director, Manpower and Personnel Laboratory,  
NPRDC (Code 06)  
San Diego, CA 92152-6800

Director, Human Factors  
& Organizational Systems Lab,  
NPRDC (Code 07)  
San Diego, CA 92152-6800

Fleet Support Office,  
NPRDC (Code 301)  
San Diego, CA 92152-6800

Library, NPRDC  
Code P201L  
San Diego, CA 92152-6800
University of Illinois/Levine

Commanding Officer, Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O’Neil, Jr.
School of Education - MPH 801
Department of Educational Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Barry Riegelhaupt
HumRRO
1100 South Washington Street
Alexandria, VA 22314

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
3108 AustinPeay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036
Lowell Schoer  
Psychological & Quantitative Foundations  
College of Education  
University of Iowa  
Iowa City, IA 52242

Dr. Mary Schratz  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. Dan Segall  
Navy Personnel R&D Center  
San Diego, CA 92152

Dr. W. Steve Sellman  
OASD(MRA&L)  
2B269 The Pentagon  
Washington, DC 20301

Dr. Kazuo Shigemasu  
7-9-24 Kugenuma-Kaigan  
Fujusawa 251  
JAPAN

Dr. William Sims  
Center for Naval Analysis  
4401 Ford Avenue  
P.O. Box 16268  
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko  
Manpower Research and Advisory Services  
Smithsonian Institution  
801 North Pitt Street  
Alexandria, VA 22314

Dr. Richard E. Snow  
Department of Psychology  
Stanford University  
Stanford, CA 94306

Dr. Richard Sorensen  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. Paul Speckman  
University of Missouri  
Department of Statistics  
Columbia, MO 65201

Dr. Judy Spray  
ACT  
P.O. Box 168  
Iowa City, IA 52243

Dr. Martha Stocking  
Educational Testing Service  
Princeton, NJ 08541

Dr. Peter Stoloff  
Center for Naval Analysis  
200 North Beauregard Street  
Alexandria, VA 22311

Dr. William Stout  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright St.  
Champaign, IL 61820

Maj. Bill Strickland  
AF/MPXOA  
4E168 Pentagon  
Washington, DC 20330

Dr. Hariharan Swaminathan  
Laboratory of Psychometric and Evaluation Research  
School of Education  
University of Massachusetts  
Amherst, MA 01003

Mr. Brad Symson  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. John Tangney  
AFOSR/NL  
Bolling AFB, DC 20332

Dr. Kikumi Tatsuoka  
CERL  
252 Engineering Research Laboratory  
Urbana, IL 61801

Dr. Maurice Tatsuoka  
220 Education Bldg  
1310 S. Sixth St.  
Champaign, IL 61820
Dr. Martin F. Wiskoff  
Navy Personnel R & D Center  
San Diego, CA 92152-6800

Mr. John H. Wolfe  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. George Wong  
Biostatistics Laboratory  
Memorial Sloan-Kettering Cancer Center  
1275 York Avenue  
New York, NY 10021

Dr. Wallace Wulfeck, III  
Navy Personnel R&D Center  
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto  
Computer-based Education Research Laboratory  
University of Illinois  
Urbana, IL 61801

Dr. Wendy Yen  
CTB/McGraw Hill  
Del Monte Research Park  
Monterey, CA 93940

Dr. Joseph L. Young  
Memory & Cognitive Processes  
National Science Foundation  
Washington, DC 20550

Dr. Anthony R. Zara  
National Council of State Boards of Nursing, Inc.  
625 North Michigan Ave.  
Suite 1544  
Chicago, IL 60611
END
Feb.
1988
DTIC