An Extension of the Two-Parameter Logistic Model to the Multidimensional Latent Space.

McKinley, Robert L.; Reckase, Mark D.
American Coll. Testing Program, Iowa City, Iowa.
ONR83-2
Aug 83
N00014-81-K0817
33p.
Reports - Research/Technical (143)

Item response theory (IRT) has proven to be a very powerful and useful measurement tool. However, most of the IRT models that have been proposed, and all of the models commonly used, require the assumption of unidimensionality, which prevents their application to a wide range of tests. The few models that have been proposed for use with multidimensional data have not been developed to the point that they can be applied in actual testing situations. The purpose of this report is to present a model for use with multidimensional data and to discuss some of its characteristics. This discussion will include information on the interpretation of the model parameters, the sufficient statistics for the model parameters, and the information function for the model. In addition, the estimation of the parameters of the model using the maximum likelihood estimation technique is also discussed. (Author/PN)
An Extension of the Two-Parameter Logistic Model to the Multidimensional Latent Space

Robert L. McKinley and Mark D. Reckase

Research Report ONR83-2
August 1983

The American College Testing Program
Resident Programs Department
Iowa City, Iowa 52243

Prepared under Contract No. N00014-81-K0817
with the Personnel and Training Research Programs
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.
**Title:** An Extension of the Two-Parameter Logistic Model to the Multidimensional Latent Space.

**Authors:** Robert L. McKinley and Mark D. Reckase

**Performing Organization Name and Address:**
The American College Testing Program
P.O. Box 168
Iowa City, IA 52243

**Controlling Office Name and Address:**
Personnel and Training Research Programs
Office of Naval Research
Arlington, VA 22217

**Abstract:**
A multidimensional extension of the two-parameter logistic latent trait model is presented and some of its characteristics are discussed. In addition, sufficient statistics for the parameters of the model are derived, as is the information function. Finally, the estimation of the parameters of the model using the maximum likelihood estimation technique is also discussed.
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>The Model and Its Characteristics.</td>
<td>1</td>
</tr>
<tr>
<td>The Model</td>
<td>1</td>
</tr>
<tr>
<td>Interpretation of the Model Parameters</td>
<td>2</td>
</tr>
<tr>
<td>Sufficient Statistics</td>
<td>5</td>
</tr>
<tr>
<td>Definition</td>
<td>5</td>
</tr>
<tr>
<td>Sufficient Statistic for the Ability Parameter</td>
<td>6</td>
</tr>
<tr>
<td>Sufficient Statistics for the Item Parameters</td>
<td>6</td>
</tr>
<tr>
<td>Information Function</td>
<td>6</td>
</tr>
<tr>
<td>Definition</td>
<td>6</td>
</tr>
<tr>
<td>Information Functions for the M2PL Model</td>
<td>8</td>
</tr>
<tr>
<td>Maximum Likelihood Estimation</td>
<td>9</td>
</tr>
<tr>
<td>Discussion</td>
<td>13</td>
</tr>
<tr>
<td>Directional Derivatives</td>
<td>13</td>
</tr>
<tr>
<td>Interpretation of the Model Parameters</td>
<td>14</td>
</tr>
<tr>
<td>Sufficient Statistics</td>
<td>16</td>
</tr>
<tr>
<td>Information Function</td>
<td>17</td>
</tr>
<tr>
<td>Maximum Likelihood Estimation</td>
<td>17</td>
</tr>
<tr>
<td>Summary</td>
<td>17</td>
</tr>
<tr>
<td>References</td>
<td>19</td>
</tr>
</tbody>
</table>
An Extension of the Two-Parameter Logistic Model to the Multidimensional Latent Space

Item response theory (IRT) has proven to be a very powerful and useful measurement tool. However, most of the IRT models that have been proposed, and all of the models commonly used, require the assumption of unidimensionality, which prevents their application to a wide range of tests. The few models that have been proposed for use with multidimensional data have not been developed to the point that they can be applied in actual testing situations. The purpose of this report is to present a model for use with multidimensional data and to discuss some of its characteristics. This discussion will include information on the interpretation of the model parameters, the sufficient statistics for the model parameters, and the information function for the model. In addition, a procedure for estimating the parameters of the model will be discussed.

The Model and Its Characteristics

The Model

The model proposed in this report is a multidimensional extension of the two-parameter logistic model. The two-parameter logistic (2PL) model, proposed by Birnbaum (1968), is given by

\[ P_i(\theta_j) = \frac{\exp(D_{ai}(\theta_j - b_i))}{1 + \exp(D_{ai}(\theta_j - b_i))}, \]  

(1)

where \( a_i \) is the discrimination parameter for item \( i \), \( b_i \) is the difficulty parameter for item \( i \), \( \theta_j \) is the ability parameter for examinee \( j \), \( P_i(\theta_j) \) is the probability of a correct response to item \( i \) by examinee \( j \), and \( D = \frac{1.7}{\theta_i} \).

The multidimensional extension of the 2PL model (M2PL), as presented by McKinley and Reckase (1982), is given by

\[ P_i(\theta_j) = \frac{\exp(d_i + a_i \theta_j)}{1 + \exp(d_i + a_i \theta_j)}, \]  

(2)

where \( a_i \) is a row vector of discrimination parameters for item \( i \), \( \theta_j \) is a column vector of ability parameters for examinee \( j \), \( P_i(\theta_j) \) is the probability of a correct response to item \( i \) by examinee \( j \), and \( d_i \) is given by

\[ d_i = \sum_{k=1}^{m} a_{ik} b_{ik}, \]  

(3)

where \( a_{ik} \) is the discrimination parameter for item \( i \) on dimension \( k \), \( b_{ik} \) is the difficulty parameter for item \( i \) on dimension \( k \), and \( m \) is the number of dimensions being modeled. The \( d_i \) term, then, is related to item difficulty,
but is not a difficulty parameter in the same sense as the $b_i$ parameter is in the unidimensional model.

**Interpretation of the Model Parameters**

The interpretation of the parameters of unidimensional IRT models is closely tied to the item characteristic curve (the regression of item score on ability). The item difficulty parameter is defined as the point on the ability scale where the point of inflection of the item characteristic curve (ICC) occurs. This is equivalent to saying the item difficulty value is the point on the ability scale where the second derivative of the ICC function is equal to zero. For the 2PL model, the second derivative is given by

$$
\frac{\delta^2 P}{\delta \theta^2} = D^2 a_i^2 Q(1 - 2P),
$$

(4)

where $P$ is the probability of a correct response to item $i$ given ability $j$, $Q = 1 - P$, and $a_i$ and $D$ are as previously defined. Setting the right hand side of (4) equal to zero yields a solution of $P = Q = 0.5$. Of course, $P = 1.0$ and $P = 0.0$ are also solutions, but these represent degenerate cases where $\theta = +\infty$ and $\theta = -\infty$, respectively. Thus, the point of inflection occurs at $P = 0.5$, which occurs where $b_i = 0_j$. The difficulty of an item for the 2PL model, then, is the point on the ability scale which yields a probability of a correct response equal to $0.5$. Figure 1 shows a typical ICC for the 2PL model. The dotted line shows the relationship among the item difficulty value, ability, and the probability of a correct response.

**Figure 1**

A Typical ICC for the 2PL Model
The item discrimination parameter is related to the slope of the ICC at the point of inflection. The slope of the ICC at the point of inflection is found by taking the first derivative of the ICC and evaluating it at the point of inflection. For the 2PL model, the first derivative is given by

\[
\frac{\delta P}{\delta \theta_j} = \alpha_j \delta \theta_j \text{,} \tag{5}
\]

where all the terms are as previously defined. It was previously found that the point of inflection for the 2PL model occurs where \( P = 0.5 \). Substituting 0.5 into (5) yields a slope at the point of inflection of \( \alpha_j/4 \).

Difficulty and discrimination are defined somewhat differently for multidimensional models. To begin with, the response function (the model) defines a multidimensional item response surface (IRS) rather than a curve. This surface may have many points of inflection, and the points of inflection may vary depending on the direction relative to the \( \theta \)-axes. Because of this, the item parameters for the M2PL model are defined in terms of directional derivatives (Kaplan, 1952).

For multidimensional models, difficulty is defined as the locus of points of inflection of the IRS for a particular direction. This is found by taking the second directional derivative of the response function, setting it equal to zero, and solving for the \( \theta \)-vector. The second directional derivative for the M2PL model is given by

\[
\frac{\delta^2 P}{\delta \theta_1^2} \cos^2 \phi_1 + \frac{\delta^2 P}{\delta \theta_1 \delta \theta_2} \cos \phi_1 \cos \phi_2 + \ldots + \frac{\delta^2 P}{\delta \theta_1 \delta \theta_m} \cos \phi_1 \cos \phi_m + \frac{\delta^2 P}{\delta \theta_2^2} \cos \phi_2 \cos \phi_1 + \ldots + \frac{\delta^2 P}{\delta \theta_2 \delta \theta_m} \cos \phi_2 \cos \phi_m + \ldots + \frac{\delta^2 P}{\delta \theta_m^2} \cos \phi_m \text{,} \tag{6}
\]

where \( \phi \) represents the vector of angles with respect to each of the \( m \) axes. Solving the derivatives in (6) and simplifying yields

\[
\frac{\delta^2 P}{\delta \theta_1^2} = PQ(1 - 2P) (a_1 \cos \phi_1 + a_2 \cos \phi_2 + \ldots + a_m \cos \phi_m)^2 \text{.} \tag{7}
\]
Setting (7) equal to zero and solving yields three solutions—\( P = 0.0 \), \( P = 0.5 \), and \( P = 1.0 \). The solutions 0.0 and 1.0 represent degenerate cases where \( \theta = \pm \infty \). \( P = 0.5 \) occurs when the exponent of the M2PL model is zero. That is \( P = 0.5 \) when

\[
d + a_1 \theta_1 + a_2 \theta_2 + \ldots + a_m \theta_m = 0. \tag{8}
\]

In the two-dimensional case this is the equation for a line.

For the M2PL model, as can be seen from the above derivations, the direction, \( \phi \), falls out of the equations. Item difficulty for the M2PL model is the same for all directions of travel. This is not necessarily the case for all multidimensional models.

Figure 2 shows a typical response surface for the M2PL model in the two-dimensional case. The dotted line indicates the line of difficulty. In the \( m \)-dimensional case (8) is the equation for a hyperplane.

---

**Figure 2**

A Typical Response Surface for the M2PL Model
For multidimensional models, item discrimination is a function of the slope of the IRS at the locus of points of inflection in a particular direction. This is obtained by taking the first directional derivative of the response function and evaluating it at the locus of points of inflection.

For the M2PL model the first directional derivative is given by

\[
\nabla \hat{P} = \frac{\delta P}{\delta \theta_1} \cos \phi_1 + \frac{\delta P}{\delta \theta_2} \cos \phi_2 + \ldots + \frac{\delta P}{\delta \theta_m} \cos \phi_m,
\]

where \( \phi \) represents the vector of angles of the direction in the \( \theta \)-space with respect to each of the \( m \) axes. For the two-dimensional case (9) is given by

\[
\nabla \hat{P} = a_1 PQ \cos \phi + a_2 PQ \sin \phi.
\]

where \( \phi \) is the angle with the \( \theta_1 \) axis. When \( \phi = 0^\circ \) (direction parallel to \( \theta_1 \) axis) the slope is \( a_1 PQ \), and when \( \phi = 90^\circ \) (parallel to \( \theta_2 \) axis) the slope is \( a_2 PQ \). In general, when the direction is parallel to the \( \theta_m \) axis, the slope is \( a_m PQ \). Since \( P = Q = 0.5 \) at the line of inflection, the slope parallel to the \( \theta_m \) axis at those points is \( a_m/4 \). In the unidimensional case \( \phi = 0^\circ \), and the slope of the ICC at the point of inflection is \( D_a/4 \).

Sufficient Statistics

**Definition** Assume that there exists some distribution that is of known form except for some unknown parameter \( \theta \), and that \( x \) represents a set of observations from that distribution. Also assume that \( S(x) \) is some statistic which is a function of \( x \). If \( S(x) \) is a sufficient statistic for \( \theta \), then it must be possible to factor the probability function of \( x \), \( P(x|\theta) \) into the form:

\[
P(x|\theta) = f[S(x)|\theta]g(x).
\]

In this form it is easy to see that \( g(x) \) is independent of \( \theta \), and so provides no information about \( \theta \). Selection of \( \theta \) to maximize the probability of \( x \) is tantamount to selecting \( \theta \) to maximize the probability of \( S(x) \).

In item response theory \( x \) is typically a response string, either by one examinee to a set of items or by a set of examinees to a single item. In this case, \( P(x|\theta) \) is the likelihood of the response string. For the M2PL model, the likelihood of an examinee's response string is given by

\[
P(x_j|\theta_j) = \prod_{i=1}^{n} P(x_{ij}|\theta_{ij})
\]

where \( x_{ij} \) is the response to item \( i \) by examinee \( j \), \( \theta_{ij} \) is the vector of abilities for examinee \( j \), \( x_j \) is the response string for examinee \( j \), and \( n \) is the number of items. The likelihood of the set of responses to an item is given by:
\[ P(\mathbf{x}_i | d_i, \mathbf{a}_i) = \prod_{j=1}^{N} P(x_{ij} | d_i, a_{ij}), \quad (13) \]

where \( P(x_{ij} | d_i, a_{ij}) \) is the probability of response \( x_{ij} \) for item \( i \), \( d_i \) and \( a_{ij} \) are the item parameters for item \( i \), \( \mathbf{x}_i \) is the vector of responses to item \( i \), and \( N \) is the number of examinees. In order for any statistic to be a sufficient statistic for a parameter of the M2PL model, it must be possible to factor the appropriate likelihood function into the form given by (11).

**Sufficient Statistic for the Ability Parameter**

For the M2PL model (12) can be factored into the form:

\[ P(\mathbf{x}_i | \theta) = \prod_{i=1}^{N} Q_i(\theta_j) \exp(\theta_j \Sigma a_{ij} x_{ij}) \exp(\Sigma d_i x_{ij}). \quad (14) \]

From (14) it can be seen that

\[ s(\mathbf{x}_i) = \sum_{i=1}^{N} a_{ij} x_{ij} \quad (15) \]

is a vector of sufficient statistics for \( \theta_j \). (For a discussion of the derivation of the sufficient statistic for ability in the unidimensional case, see Lord and Novick, 1968, chapter 18).

**Sufficient Statistics for the Item Parameters**

For the item parameters of the M2PL model, (13) can be factored into the form:

\[ P(\mathbf{x}_i | d_i, \mathbf{a}_i) = \prod_{j=1}^{N} P(x_{ij} | d_i, a_{ij}) \exp(\Sigma d_i x_{ij}) \exp(\Sigma a_{ij} x_{ij}). \quad (16) \]

From (16) it can be seen that

\[ s_d(\mathbf{x}_i) = \sum_{j=1}^{N} x_{ij} \quad (17) \]

is a sufficient statistic for the \( d \)-parameter, and

\[ s_a(\mathbf{x}_i) = \sum_{j=1}^{N} \theta_j x_{ij} \quad (18) \]

is a vector of sufficient statistics for the \( a \)-parameter.

**Information Function**

**Definition**

In item response theory the precision of estimates based on a given scoring formula are generally described in terms of the information function of the scoring formula. The information function of a particular
scoring formula, as given by Lord and Novick (1968), is given by
\[ I[\theta, s(x)] = \frac{1}{\sigma^2 [s(X), \theta]} \left\{ \frac{\delta^2}{\delta \theta} \mathbb{E} \left[ s(X) | \theta \right] \right\}^2, \tag{19} \]

where \( s(x) \) represents a given scoring formula for the model of interest, \( \sigma^2 [s(X), \theta] \) is the variance of the scoring formula, and the derivative \( \frac{\delta}{\delta \theta} \mathbb{E} [s(X) | \theta] \) specifies how the mean of the scoring formula changes as \( \theta \) changes.

If \( s(x) \) takes the form
\[ s(x) = \sum_{i=1}^{n} w_i x_i, \tag{20} \]
where \( w_i \) is a positive number, then the expected value \( \mathbb{E} [s(X) | \theta] \) is given by
\[ \mathbb{E} \left[ s(X) | \theta \right] = \sum_{i=1}^{n} w_i \mathbb{P}_i(\theta), \tag{21} \]
and the variance of the scoring formula is given by
\[ \sigma^2 [s(X), \theta] = \sum_{i=1}^{n} w_i^2 \mathbb{P}_i(\theta) \mathbb{Q}_i(\theta). \tag{22} \]

(For a discussion of these derivations, see Lord and Novick, 1968). Substituting (21) and (22) into (19) yields
\[ I[\theta, s(x)] = \left[ \sum_{i=1}^{n} w_i^2 \mathbb{P}_i(\theta) \mathbb{Q}_i(\theta) \right]^{-1} \left[ \sum_{i=1}^{n} w_i \mathbb{P}_i(\theta) \right]^2, \tag{23} \]
where \( \mathbb{P}_i(\theta) = \frac{\partial \mathbb{P}_i(\theta)}{\partial \theta}. \) For a single item (23) takes the form
\[ I[\theta, s(x)] = \mathbb{P}_i(\theta)^2 / \mathbb{P}_i(\theta) \mathbb{Q}_i(\theta), \tag{24} \]
which is the item information function. If (24) is written in terms of the response \( x_i \), rather than the scoring formula \( s(x) \), the same result is obtained. That is, \( I[\theta, x_i] = I[\theta, s(x_i)] \). Lord and Novick (1968) have shown that, unless \( s(x) \) represents the locally best weights at \( \theta \), \( I[\theta, s(x)] < \sum I[\theta, x_i] \). That is, the sum of the item information functions, which is independent of the the scoring formula, represents an upper bound on each and all information functions obtained using different scoring formulas. The sum of the item information functions is called the test information function, and is given by
Information Functions for the M2PL Model For the unidimensional 2PL model, given by (1), the item information function is given by

\[ I(\theta, x_i) = \frac{D^2 a_i^2 P_i(\theta) Q_i(\theta)}{1} \]  

(25)

Test information for the unidimensional 2PL model is given by

\[ I(\theta) = \sum_{i=1}^{n} \frac{D^2 a_i^2 P_i(\theta) Q_i(\theta)}{1} \]  

(26)

As was the case for discrimination, information for the M2PL model varies depending on the direction relative to the \( \theta \)-axes. Therefore, item and test information for the M2PL model are defined using the first directional derivative of the response function, which is given by (9). Item information for the M2PL model is given by

\[ I(\theta, x_i) = a_1^2 PQ \cos^2 \phi_1 + a_2^2 PQ \cos^2 \phi_2 + \ldots + a_m^2 PQ \cos^2 \phi_m + \]

\[ 2a_1a_2 PQ \cos \phi_1 \cos \phi_2 + \ldots + 2a_1a_m PQ \cos \phi_1 \cos \phi_m + \]

\[ \ldots \]

\[ 2a_{(m-1)}a_m PQ \cos \phi_{(m-1)} \cos \phi_m. \]  

(28)

For the two dimensional case, this simplifies to

\[ I(\theta, x_i) = PQ(a_1 \cos \phi + a_2 \sin \phi)^2. \]  

(29)
Note that when the direction of travel is parallel to the \( \theta \)-axis (\( \phi = 0^\circ \)), item information is given by \( a_{12}PQ \). When only \( \theta_2 \) is of interest (\( \phi = 90^\circ \)), item information is given by \( a_{22}PQ \). If the two dimensions are weighted equally (\( \phi = 45^\circ \)), item information is given by \( 0.5(a_{12}PQ + 2a_{11}a_{22}PQ + a_{22}PQ) \). Figures 3, 4, and 5 show typical item information surfaces for \( \phi = 0^\circ \), 45\( ^\circ \), and 90\( ^\circ \), respectively. Note that these are not the same surface seen from different angles. They are different surfaces, all for the same item, obtained by varying the direction with respect to the \( \theta \)-axes. As can be seen, they are quite different. Test information for the M2PL model is simply the sum of (29) over all of the items. Figures 6, 7, and 8 show typical test information surfaces for \( \phi = 0^\circ \), 45\( ^\circ \), and 90\( ^\circ \), respectively. Again, the three surfaces are quite different, indicating that the test gives different amounts of information that are concentrated at different places in the \( \theta \)-space when different weighted composites of ability are of interest.

**Maximum Likelihood Estimation**

Maximum likelihood estimation of the parameters of the M2PL model is relatively straightforward. The likelihood of a response matrix for the M2PL model (or for any latent trait model) is given by

\[
L = \prod_{i=1}^{n} \prod_{j=1}^{N} P(x_{ij})
\]

where all the terms have been previously defined. For an examinee's response string, the likelihood is given by (12), and the likelihood of a response string for an item is given by (13). The first derivative of the log of the likelihood given in (12) is given by:

\[
\frac{\delta \log L}{\delta \theta} = \sum_{i=1}^{n} a_{ii} x_{ij} - \sum_{i=1}^{n} a_{ii} P_{ij}
\]

and the first derivative of the log of the likelihood given in (13) is given by

\[
\frac{\delta \log L}{\delta d_i} = \sum_{j=1}^{N} x_{ij}
\]

for the difficulty parameter, and

\[
\frac{\delta \log L}{\delta a_i} = \sum_{j=1}^{N} \theta_j x_{ij} - \sum_{j=1}^{N} \theta_j P_{ij}
\]

for the discrimination parameter.

The estimation of ability using maximum likelihood techniques simply involves setting (31) equal to zero and solving for \( \theta_j \). Of course, since this involves solving simultaneous nonlinear equations, some type of iterative procedure is generally required. The estimation of item parameters involves setting (32) and (33) equal to zero and solving for \( d_i \) and \( a_i \).
Figure 3

An Item Information Surface

$\theta = 0^\circ$

Figure 4

An Item Information Surface

$\theta = 45^\circ$
Figure 5

An Item Information Surface

\[ \phi = 90^\circ \]

Figure 6

A Test Information Surface

\[ \phi = 0^\circ \]
Figure 7

A Test Information Surface

$\phi = 55^\circ$

Figure 8

A Test Information Surface

$\phi = 90^\circ$
respectively. Again, the solution of simultaneous nonlinear equations requires an iterative procedure. McKinley and Reckase (1983) describe a procedure for the simultaneous estimation of the item and person parameters of the M2PL model using a Newton-Raphson procedure for solving the simultaneous nonlinear equations. A computer program is available.

Discussion

Although IRT has gained popularity over the last few years, applications of IRT models have been limited to tests for which the assumption of unidimensionality is at least defensible. There have been a few IRT models proposed for use with multidimensional data (see McKinley and Reckase, 1982, for a summary), but there have been few successful attempts at their application. Use of these models has been limited due to the absence of practical algorithms for parameter estimation, and, at least in part, because the models are not well understood.

McKinley and Reckase (1982) have proposed a model, the M2PL model, for use with multidimensional data, and they have developed a program for the estimation of the parameters of the model (McKinley and Reckase, 1983). The purpose of this report is to provide information necessary for the understanding and use of the M2PL model.

Many of the characteristics of the M2PL model are not straightforward extensions from the unidimensional case. Rather, the unidimensional case is a special case of the multidimensional model in which much of the richness and complexity of the model is not evident. Because of this, some of the characteristics of the model described in this report may be somewhat difficult to grasp. In order to aid in the understanding of these characteristics, they will now be discussed in some depth. An attempt will be made in each case to describe how the information provided relates to real-world applications. The discussion will begin with the interpretation of the model parameters, and will include the sufficient statistics, information functions, and parameter estimation. Before beginning the discussion of the characteristics of the M2PL model, however, a brief discussion of directional derivatives will be presented, since directional derivatives are so important to the understanding of multidimensional IRT models.

Directional Derivatives

One of the most interesting and complex aspects of the multidimensional IRT models which is lost when the unidimensional case is discussed is the notion of directional derivatives. In the unidimensional case the only direction ever discussed is parallel to the $\theta$-axis ($\phi = 0^\circ$), in which case all the trigonometric terms so evident in the material presented above are absent—they always equal 1.0 or 0.0 and therefore drop out of the equations.

Directional derivatives are necessary in the multidimensional case because the derivatives of the response function vary depending on the direction taken relative to the $\theta$-axes. The first derivative of a function gives the slope of the function at any given point. The second derivative gives the rate at which the slope is changing at a particular point. The point of maximum slope is where the slope stops increasing and starts decreasing. At the point where that change occurs, the change in slope
crosses from being positive (increasing) to negative (decreasing). Thus, at that point the change in slope is neither positive nor negative, but rather is zero. Since the second derivative gives the rate of change of slope, the point of maximum slope is where the second derivative is zero. In the unidimensional case, this has a straightforward application in determining item difficulty and discrimination, as illustrated in Figure 1. The point where the dotted line crosses the ICC is the point of maximum slope and minimum (zero) change in slope.

In Figure 2, it can easily be seen that there is no one point on the surface where the slope is at a maximum. Moreover, for any one point, the slope varies depending on the direction. Consider the point on the surface where \( \theta_1 = 0.0 \) and \( \theta_2 = -2.5 \). This point is indicated on the surface by an x. Moving along the \( \theta_2 = -2.5 \) line parallel to the \( \theta_1 \) axis, the surface is rising fairly rapidly at the point indicated. However, moving along the \( \theta_1 = 0 \) line parallel to the \( \theta_2 \)-axis, the surface is still relatively flat and is rising slowly. Clearly the slope of the surface is different depending on the direction of travel. The same is true of the change in slope. Because of this, when taking derivatives of a multidimensional response function, it is necessary to consider the direction. Directional derivatives are a way of doing this. The actual interpretation of the derivatives in different directions will be addressed in the next section, since it is closely related to the interpretation of the model parameters.

Interpretation of the Model

A straightforward extension of item difficulty from the unidimensional to the multidimensional case would seem to lead to the conclusion that difficulty in the multidimensional case ought to be a vector of \( b \)-parameters, with one \( b \) for each dimension. In Figure 1 the \( b \)-parameter is the point on the \( \theta \)-scale below the point of inflection. It represents the point on the \( \theta \)-scale where the item best discriminates between high and low ability. At the point represented by the \( b \)-parameter, a very small change in ability corresponds to a large change in the probability of a correct response. Nowhere on the \( \theta \)-scale does an equal change in ability result in as large a change in the probability of a correct response. Thus, in the unidimensional case, the item difficulty parameter indicates the point on the ability scale at which the item does the best job of discriminating between different levels of ability.

On the surface, it seems logical to conclude that if there are two dimensions, there should be two \( b \)-parameters. One \( b \)-parameter should indicate the point of maximum discrimination on one dimension, while the other \( b \)-parameters indicate the point of maximum discrimination on the other dimension. Figure 2, however, clearly illustrates that this is inadequate.

As can be seen in Figure 2, the two ability dimensions do not act independently. It is the combination of ability on the two dimensions which determines the probability of a correct response. An examinee with \( \theta_2 = 2.0 \) clearly has a higher ability on that dimension than an examinee with \( \theta_2 = -2.0 \). However, if the second examinee has \( \theta_1 = 3.0 \), while the first examinee has \( \theta_1 = -3.0 \), the second examinee has a much higher probability of a correct response to the item described by the IRS. Clearly, then, considering the two dimensions separately does not contribute greatly
to discriminating between examinees who have different probabilities of a correct response. This is reflected in the fact that the item difficulty for Figure 2 is a line which is not parallel to either axis.

This has important implications for test construction and analysis using the M2PL model. It is common practice, for instance, to order items on a test by difficulty, or to construct a test having a specified distribution of item difficulty. In the unidimensional case this is done using the b-parameter. Clearly in the multidimensional case it is more complicated. An item having a smaller d-parameter than a second item is only uniformly more difficult than the second item if their difficulty functions are parallel. If the difficulty functions intersect, then item I is more difficult than item 2 in some regions of the $\theta$-plane, while item 2 is more difficult in other regions.

This would seem to indicate that it is only reasonable, in the multidimensional case, to talk about ordering items on difficulty or obtaining a specified distribution of difficulty if all the items to be considered have parallel lines of difficulty. Of course, in the $m$-dimensional case the items would all have to have parallel $(m-1)$-dimensional hyperplanes.

In order to determine whether two items have parallel lines of difficulty in the two-dimensional case, first determine the form of the difficulty line. The equation for the line of difficulty is given by (8). The two lines are parallel only if the slopes of the lines are equal. Putting (8) into a slope-intercept form yields

$$\theta_{j2} = -\frac{a_{11}}{a_{12}} \theta_{j1} + \frac{d_1}{a_{12}},$$

where $a_{11}$ is the item discrimination parameter for item $i$ for dimension 1, $a_{12}$ is for dimension 2, $\theta_{j1}$ is the ability parameter for examinee $j$ for dimension 1, and $\theta_{j2}$ is the ability parameter for dimension 2. If item 2 is denoted by a prime (′), then the lines of difficulty for items 1 and 2 are parallel only if

$$\frac{a_{11}}{a_{12}} = \frac{a_{11}'}{a_{12}'}.$$ (35)

If all items meet the condition set out in (35), then they can be ordered by difficulty, by simply ordering them by their $d$-parameters.

The ordering of items on difficulty implies that there is some underlying variable being measured that has some correspondence to the criterion used for the ordering. In this case there is some composite of the $\theta$s which corresponds to the difficulty continuum formed by the items having parallel lines of difficulty. The composite is determined by the orientation of the lines of difficulty.

The extension of item discrimination to the multidimensional case is even more complex than the extension of item difficulty. Unlike difficulty, the concept of item discrimination in the multidimensional case includes a consideration of direction—the angles indicating the direction do not fall
out of the equations. Although the slope of the IRS shown in Figure 2 is constant all along the line of difficulty for a given direction, it varies with different directions.

The need to consider direction has important implications for both test construction and test analysis. It is not enough in test construction, for instance, to merely select the item with the highest discrimination from among the available items. One item is uniformly more discriminating than a second item only if the slope of its IRS at the points of inflection is greater than the slope of the IRS for the second item for all directions. If item 1 has the higher \( a \)-value on one dimension, but a lower \( a \)-value on another dimension, there may be directions for which the slope of the IRS at the points of inflection will be greater for item 2. For example, consider the case where item 1 has discrimination parameters \( a = (1.0, 0.5) \) and item 2 has discrimination parameters \( a = (0.5, 1.0) \). When \( \phi \) in (10) is 30 degrees, the slope of the IRS at the points of inflection is 0.279 for item 1 and 0.233 for item 2. When \( \phi \) is 60 degrees, the slope for item 1 is 0.233, while the slope for item 2 is 0.279. At \( \phi = 45 \) degrees, both items have a slope of 0.265 at the points of inflection.

It seems to follow from the above discussion and example that, in interpreting item discrimination in the multidimensional case, the particular composite of abilities of interest must be considered. The composite might be specified a priori, as in test construction, or discovered by post administration analyses.

**Sufficient Statistics**

The notion of a sufficient statistic is not a simple one to grasp. Essentially, a statistic \( t \) is a sufficient statistic for the parameter \( \theta \) if it contains all the information in the sample data about \( \theta \). For example, the number-correct score for an item provides all the information in the response data about the \( d \)-parameter. For the \( a \)-parameter for a particular dimension, a sufficient statistic is provided by a weighted sum of the item responses. The response of each examinee to the item of interest is weighted by the examinee's ability on the dimension of interest. Thus, a correct response to an item by an examinee of high ability \( (\theta > 0.0) \) adds to the value of the statistic, while a correct response by an examinee of low ability \( (\theta < 0.0) \) decreases the value.

For ability, a sufficient statistic is provided by a weighted sum of an examinee's responses to all the items. The weighting factor is the discrimination parameter for the dimension of interest. Thus, a correct response to a highly discriminating item adds more to the statistic than a correct response to an item of low discriminating power.

Although the availability of sufficient statistics for the parameters of the M2PL model is important from an estimation standpoint, it should be pointed out that, with the exception of the \( d \)-parameter, the sufficient statistics described above are not observable. While the number-correct score of an item can be observed, the \( a \)-parameter of an item, and thus the sum of item responses weighted by discrimination parameters, is not observable. This complicates estimation somewhat by requiring that provisional estimates of some parameters be provided during the estimation of the remaining parameters. Solutions, then, are obtained by a series of approximations by
varying from one step to the next which parameters are estimated. In each step the provisional estimates for the parameters not being estimated are the most recent estimates of those parameters.

**Information Function**

Item information in the multidimensional case is like item discrimination in that the information yielded by an item for a particular $\theta$ varies with the direction of travel. This has important implications for such applications as adaptive testing, in which items may be selected for administration on the basis of item information. As was the case with item discrimination, the interpretation and use of item information requires the consideration of the particular composite of abilities which is of interest.

**Maximum Likelihood Estimation**

Estimation of the parameters of the M2PL model is surprisingly straightforward. Implementation of the procedure described earlier in this report is not particularly difficult. However, it is rather expensive. One serious limitation of the procedure described is that there is no way to determine in advance how many dimensions should be included. The procedure is too expensive to allow successive runs for an increasing number of dimensions until a satisfactory solution is obtained. It is clear that, if the M2PL model is to be used, more work is needed in this area.

More work also needs to be done to determine sample size requirements for estimation. Some guidelines are needed for determining the maximum number of items and subjects required for good estimation.

**Summary**

Item response theory has become an increasingly popular area for research and application in recent years. Areas where item response theory has been applied include test scoring (Woodcock, 1974), criterion-referenced measurement (Hambleton, Swaminathan, Cook, Eignor, and Gifford, 1978), test equating (Marco, 1977; Rentz and Bashaw, 1977), adaptive testing (McKinley and Reckase, 1980), and mastery testing (Patience and Reckase, 1978). While many of these applications have been successful, one unsolved problem is repeatedly encountered—most IRT models assume unidimensionality. As a result, applications of these models have been limited to areas in which the tests used measure predominantly one factor (or can be sorted into subtests which measure predominantly one factor). When the assumption of unidimensionality is not met, most IRT models are inappropriate.

The purpose of this report is to present an IRT model that does not require unidimensional tests. With such a model the great power of item response theory as a measurement tool can be applied for many of the purposes for which unidimensional models are employed, without the limitation on what kinds of tests are involved (i.e., the dimensionality of the tests). Of course, much more work is needed before the model can be employed in real testing situations. Procedures for the use of the model for different
applications must be worked out in greater detail, and limitations on the practicality of the estimation procedures must be overcome. The information provided in this report provides a firm foundation for future work in these areas.
REFERENCES


McKinley, R.L. and Reckase, M.D. A successful application of latent trait theory to tailored achievement testing (Research Report 80-1). Columbia, MO: University of Missouri, Department of Educational Psychology, February 1980.


Air Force

1 Lt. Col. James E. Watson
HQ USAF/MFXO
The Pentagon
Washington, DC 20334

1 Jr. John Welch
AFHPC
Randolph AFB, TX

1 Lt. Cra
Jim 2S E.
Wagon
Ill OSIF/MPXON
Ill. Pontasgon
Washington, Or. 20310

Department of Defense

12 Defense Technical Information Center
Cameron Station, Blg 5
Alexandria, VA 22314
Attn: TC

1 Dr. Craig L. Fields
Advanced Research Projects Agency
1800 Wilson Blvd.
Arlington, VA 22209

1 Jerry Lezama
ARPA
Ven: WA010-P
Fort Shrelin, IL 60117

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defense for Research & Engineering
Room 10129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sullivan
Office of the Assistant Secretary of Defense (DRA & L)
24169 The Pentagon
Washington, DC 20301

1 Major Jack Thomas
DARPA
1400 Wilson Blvd.
Arlington, VA 22209
Civilian Agencies

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1201 19th Street NW
Washington, DC 20201

1 Dr. Veronica W. Oury
Personnel R&D Center
Office of Personnel Management
1901 E Street NW
Washington, DC 20415

1 Mr. Thomas A. Warm
U.S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

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

Private Sector

1 Dr. James Algina
University of Florida
Gainesville, FL 326

1 Dr. Erling R. Anderson
Department of Statistics
Studierstraede 6
1456 Copenhagen
DENMARK

1 Psychological Research Unit
NW-1-44 Suite: Librarian
Northwestern Illinois
Twin Cities 2001
AUSTRALIA

1 Dr. Isaac Bajir
Educational Testing Service
Princeton, NJ 08540

1 Dr. Noam Nisan
School of Education
Tel Aviv University
Tel Aviv, 69978 Israel

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Robert Brown
American College Testing Program
P. O. Box 168
Iowa City, IA 52243

1 Dr. Ernest R. Caffette
777 Stokely
University of Tennessee
Knoxville, TN 37916

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90037
Private Sector

1 Dr. Douglas M. Jones  
Advanced Statistical Technologies  
Corporation  
10 Trivalgar Court  
Lawrenceville, NJ 08648

1 Professor John A. Knits  
Department of Psychology  
The University of New South Wales  
NSW 2052  
AUSTRALIA

1 Dr. William Knob  
University of Texas-Austin  
Measurement and Evaluation Center  
Austin, TX 78713

1 Dr. Alain Lesgold  
Learning R&D Center  
University of Pittsburgh  
1919 O'Hara Street  
Pittsburgh, PA 15260

1 Dr. Michael Levin  
Department of Educational Psychology  
210 Education Bldg.  
University of Illinois  
Champaign, IL 61820

1 Dr. Charles Lewa  
Faculteit Sociale Wetenschappen  
Rijksuniversiteit Groningen  
Oude Boteringstraat 21  
9712GC Groningen  
Netherlands

1 Dr. Robert Linn  
College of Education  
University of Illinois  
Urbana, IL 61801

1 Mr. Phillip Livingston  
Systems and Applied Science Corporatio  
6311 Kenilworth Avenue  
Riverdale, MD 20820

1 Dr. Robert Lockman  
Center for Naval Analysis  
200 North Bouregard St.  
Alexandria, VA 22314

Private Sector

1 Dr. Frederic W. Lord  
Educational Testing Service  
Princeton, NJ 08541

1 Dr. James Lumsden  
Department of Psychology  
University of Western Australia  
N.1.A. 6009  
AUSTRALIA

1 Dr. Gary Yonin  
Test III  
Educational Testing Service  
Princeton, NJ 08541

1 Dr. Scott Westell  
Department of Psychology  
University of Notre Dame  
Notre Dame, IN 46556

1 Dr. Samuel T. Hino  
Loyola University of Chicago  
920 North Michigan Avenue  
Chicago, IL 60611

1 Mr. Robert McKinley  
American College Testing Programs  
P.O. Box 168  
Iowa City, IA 52243

1 Dr. Barbara Mims  
Human Resources Research Organization  
703 North Washington  
Alexandria, VA 22314

1 Dr. Robert Mislevy  
711 Illinois Street  
Carroll, IA 50414

1 Dr. Allen Muir  
Behavioral Technology Laboratories  
1865 Elliott Ave., Fourth Floor  
Redondo Beach, CA 90277

1 Dr. W. Alan McConaughy  
Department of Psychology  
University of Oklahoma  
Oklahoma City, OK 73109

1 Dr. Melvin R. Novick  
156 Lindquist Center for Measurement  
University of Iowa  
Iowa City, IA 52242
Private Sector

Dr. James Olson
WICAT, Inc.
1875 South State Street
Drum, UT 84057

Wayne M. Patience
American Council on Education
GED Testing Services, Suite 29
One Dupont Circle, NW
Washington, DC 20036

Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

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

Dr. Thomas Reynolds
University of Texas-Dallas
Marketing Department
P.O. Box 588
Richardson, TX 75080

Dr. Lawrence Rutner
433 Elm Avenue
Takoma Park, MD 20912

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

Prof. Fumiko Samejima
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

Dr. Frank L. Schmidt
Department of Psychology
Bldg. GC
George Washington University
Washington, DC 20052

Dr. Walter Schneider
Psychology Department
605 E. Daniel
Champaign, IL 61820

Private Sector

Lowell School
Psychological & Quantitative Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Robert J. Seidel
INSTRUCTIONAL TECHNOLOGY GROUP
110 N. Washington St.
ALEXANDRIA, VA 22314

Dr. Kiyuz Shigehara
University of Tokyo
Department of Educational Psychology
Kawachi, Saito 940
JAPAN

Dr. Edwin Shirkley
Department of Psychology
University of Central Florida
Orlando, FL 32816

Dr. William Sims
Center for Naval Analyses
212 North Beauregard Street
Alexandria, VA 22311

Dr. H. Wallace Staniko
Program Director
Handower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 114, Yale Station
New Haven, CT 06520

Dr. P-Lor Stoloff
Center for Naval Analyses
212 North Beauregard Street
Alexandria, VA 22311

Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801
Private Sector

1 Dr. Kihito Fumio
Laboratory of Psychometric and Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsukuri
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Kurio Tatsukuri
224 Education Bldg
1316 S. Sixth St.
Champaign, IL 61820

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66045

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. J. Uhler
Uhler Consultants
4258 San Vicente Drive
Encino, CA 91436

1 Dr. V. N. R. Uppaluri
Union Carbide Corporation
Nuclear Division
P. O. Box Y
Oak Ridge, TN 37830

1 Dr. David Vile
Assessment Systems Corporation
2211 University Avenue, Suite 310
St. Paul, MN 55114

1 Dr. Howard Witmer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 Dr. Michael T. Waller
Department of Educational Psychology
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

Private Sector

1 Dr. Brian Waters
HunRRO
333 North Washington
Alexandria, VA 22314

1 Dr. David J. Weiss
4667 Elliott Hill
University of Michigan
15 S. River Rd.
Ann Arbor, MI 48105

1 Dr. Raul R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90031

1 Wolfgang Wildgruber
Streitkrafeanst
Box 29 59 33
D-5001 Bonn 2
WEST GERMANY

1 Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

1 Dr. Wendy Yun
CIBA/Grurray Hill
D-1 Institute Research Park
Monterey, CA 93940