DOCUMENT RESUME

ED 264 265 TM 850 750

AUTHOR	Reckase, Mark D.; McKinley, Robert L.
TITLE	Some Latent Trait Theory in a Multidimensional Latent Space.
PUB DATE	Jul 82
NOTE	28p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Estimation (Mathematics); Higher Education; *Latent Trait Theory; *Mathematical Models; *Multidimensional Scaling; *Predictive Measurement; Psychometrics; Simulation; Tables (Data); Test Items
IDENTIFIERS	ACT Assessment; *Parametric Analysis; Rasch Model; *Unidimensional Scaling

ABSTRACT

This paper reviews the existing multidimensional item response theory (IRT) models and demonstrates how one of the models can be applied to estimation of abilities from a test measuring more than one dimension. The purposes of this paper were threefold. First, the fundamental concepts required when considering multidimensional models for the interaction of a person and a test item were defined. These concepts included the multidimensional latent space, the item difficulty function, and the item discrimination function. These definitions were conceived as multidimensional generalizations of similar concepts in unidimensional IRT models. Second, six existing multidimensional models were reviewed and, on the basis of their similarities, were classified into three general categories. The characteristics of these categories were described, and the general Rasch model was selected for further study on the basis of ease of parameter estimation. Third, estimation procedures for the parameters of the general Rasch model were described and applied to a set of simulation data that had been generated according to a two-dimensional special case of the model. The results indicated that a very close correspondence had been obtained between the estimated item parameters and those used to generate the simulation data. (PN)

DOCUMENT RESUME

ED 264 265 TM 850 750

AUTHOR Reckase, Mark D.; McKinley, Robert L.
TITLE Some Latent Trait Theory in a Multidimensional Latent Space.
PUB DATE Jul 82
NOTE 28p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Estimation (Mathematics); Higher Education; *Latent Trait Theory; *Mathematical Models; *Multidimensional Scaling; *Predictive Measurement; Psychometrics; Simulation; Tables (Data); Test Items
IDENTIFIERS ACT Assessment; *Parametric Analysis; Rasch Model; *Unidimensional Scaling

ABSTRACT

This paper reviews the existing multidimensional item response theory (IRT) models and demonstrates how one of the models can be applied to estimation of abilities from a test measuring more than one dimension. The purposes of this paper were threefold. First, the fundamental concepts required when considering multidimensional models for the interaction of a person and a test item were defined. These concepts included the multidimensional latent space, the item difficulty function, and the item discrimination function. These definitions were conceived as multidimensional generalizations of similar concepts in unidimensional IRT models. Second, six existing multidimensional models were reviewed and, on the basis of their similarities, were classified into three general categories. The characteristics of these categories were described, and the general Rasch model was selected for further study on the basis of ease of parameter estimation. Third, estimation procedures for the parameters of the general Rasch model were described and applied to a set of simulation data that had been generated according to a two-dimensional special case of the model. The results indicated that a very close correspondence had been obtained between the estimated item parameters and those used to generate the simulation data. (PN)

SOME LATENT TRAIT THEORY IN A MULTIDIMENSIONAL LATENT SPACE

MARK D. RECKASE AND ROBERT L. MCKINLEY
THE AMERICAN COLLEGE TESTING PROGRAM

Measuring instruments that are used to determine an individual's level of performance on a psychological or educational trait are seldom truly unidimensional. Certainly, tests based on number series or vocabulary knowledge may approximate unidimensional measures, but even these narrowly focused tests usually measure more than one trait (see, e.g., Holzman, Glaser, & Pellegrino, 1980).

Alternatively, many tests are purposely designed to measure more than one trait. The English Usage Test from the ACT Assessment battery, for example, measures skills in punctuation, grammar, sentence structure, diction and style, and logic and organization (American College Testing Program, 1980). These topics have been included in the test in order to assess more thoroughly the skills acquired in high school English than could be obtained from a unitary measure. There is also a statistical motivation for constructing a test that measures more than one dimension. To maximize a test's ability to predict a criterion measure, the test should have items that have a high correlation with the criterion but low intercorrelations among themselves (Lord & Novick, 1968). Following this selection rule results in a test that measures many dimensions.

The fact that measuring devices seldom measure single dimensions has serious consequences for the application of item response theory (IRT) to test data. A basic assumption of most of the IRT models currently being applied is that the measuring instrument measures a single trait (Lord, 1980; Lord & Novick, 1968). To the extent that this assumption is violated, these IRT models may not be appropriate. Since tests seldom measure single dimensions, the unidimensional IRT models are only applicable if they can be shown to be robust to the violation of the unidimensionality assumption or if the items in a test can be sorted into subtests that measure a single dimension.

The issue of the robustness of two IRT models--the 1-parameter and 3-parameter logistic models--to violations of the unidimensionality assumption has been addressed by Reckase (1977). He found that even when the proportion of variance accounted for by the dominant dimension was as low as 20%, the two models still resulted in reasonable ability estimates. However, since these estimates were of the dominant dimension, much information was lost about other traits being measured by the other dimensions in the tests. On the other hand, if the measuring instrument in question measured several traits with equal emphasis, the meaning of the ability estimates was difficult to define. Thus, although the models do seem to be somewhat robust to violations of the unidimensionality assumption, it is at the cost of lost information or poorly defined traits. It would seem that a better approach would be to estimate the ability on each dimension separately.

Two different alternatives exist for obtaining estimates of the traits measured by a test when more than one trait is being measured. First, the items in the test may be subdivided into groups of items that are sensitive to differences on one of the dimensions. This procedure breaks down the test into a series of unidimensional subtests. Unfortunately, no procedure exists that adequately performs this function when dichotomously scored test items are used (Reckase, 1981). Factor analysis is the procedure most commonly used for sorting items, but factor analysis suffers from several problems due to the choice of the correlation coefficient, the effects of guessing, and the determination of the number of factors (Kim & Mueller, 1978). Therefore, in many cases the formation of unidimensional subsets of items is not a reasonable approach.

The second possible approach for obtaining estimates of the abilities on each dimension is to develop a multidimensional model of performance that relates dichotomous item responses to the magnitude of ability on each trait. Several models of this type have been presented in the IRT literature (Bock & Aitkin, 1981; Mulaik, 1972; Rasch, 1961; Samejima, 1974; Whitely, 1980), but little work has been done using these models in an applied testing setting. In fact, little research has been done to determine the characteristics of these models or the properties of the ability estimates obtained through their use.

From this discussion it should be evident that obtaining estimates of trait levels from a test that measures more than one trait is a difficult problem. Traditional models such as factor analysis and nonmetric multidimensional scaling are not well suited for use with dichotomously scored test items, and most of the IRT models that are designed for use with dichotomous test data assume a unidimensional test. The use of multidimensional IRT models may be the solution to this problem; however, little work has been done to demonstrate their usefulness. The purpose of this paper is to review the existing multidimensional IRT models and to show how one of the models can be applied to the estimation of abilities from a test measuring more than one dimension.

Definition of the Problem

Most of the IRT models currently in use assume that the test being analyzed measures a unidimensional latent trait. This means that all persons having the same amount of the trait, θ, should have the same probability of a correct response to a dichotomously scored item. If individuals with the same level of the single trait have different probabilities of a correct response to a test item, this implies that at least one other trait is involved in responding to the item. If only two dimensions are required in the solution of the item, then all persons that have the same values on these two dimensions, $[\theta_1, \theta_2]$, should have the same probability of a correct response. Again, if the examinees have different probabilities of a correct response, at least one more dimension is indicated. Once the number of dimensions, $\underline{n}$, is determined that results in a constant probability of a correct response for all persons with the same set of abilities, $\theta_1, \theta_2, \ldots, \theta_n$, the size of the complete latent space has been defined. This concept is discussed in more detail by Lord and Novick (1968).

Note that this method of defining the size of the complete latent space emphasizes the ability dimensions of the examinees while treating the test item as a constant stimulus. No information is given concerning the characteristics of the items. In order to determine the characteristics of the test items, critical features of the surface describing the relationship between the probability of a correct response and a person's position in the θ space must be defined. Two such features that are typically used in IRT are (1) the difficulty of the item (location of the point of inflection of the item characteristic curve, or ICC) and (2) the discriminating power of the item (related to the slope of the ICC at the point of inflection). If the relationship between the probability of a correct response and a person's location in the θ space can be described by a sufficiently well-behaved mathematical function (e.g., the logistic function), the concept of difficulty and discrimination can be generalized to items that measure more than one dimension in the complete latent space.

Suppose that the relationship between the probability of a correct response to a dichotomously scored test item and a person's location in the θ space is given by a function that is monotonically increasing for all θ dimensions and is asymptotic to 0 and 1 as each $\theta \to -\infty$ and $\theta \to \infty$, respectively. That is,

$$f(\theta_{ij}) < f(\theta_{ik}) \qquad i = 1, \ldots, n , \tag{1}$$

if $\theta_{ij} < \theta_{ik}$ for all i, j, and k, where i indicates the dimension and j and k indicate the person, and

$f(\theta_{ij}) \to 0$

as $\theta_{ij} \to -\infty$,

$f(\theta_{ij}) \to 1$

as $\theta_{ij} \to \infty$, and

$1 > f(\theta_{ij}) > 0$ for all i, j .

Then, the difficulty of the item can be defined as the values of $\underset{\sim}{\theta}$ for which

$$\frac{d^2 f(\underset{\sim}{\theta})}{d\underset{\sim}{\theta}^2} = 0 , \tag{2}$$

if certain regularity conditions hold. This is the multivariate equivalent to the point of inflection of the univariate ICC.

For some functions, $f(\underset{\sim}{\theta})$, the second derivative will be undefined; and for others, Equation 2 will yield multiple solutions. However, for a class of models based on the logistic function, Equation 2 gives a solution that defines a difficulty function rather than a difficulty value for an item. This function is the locus of points in the $\underset{\sim}{\theta}$ space that yields a .5 probability of a correct response to the item. An example of the difficulty function for an item that measures two dimensions may help clarify this concept.

Suppose that the complete latent space is defined by two dimensions, θ_1 and θ_2, and that the relationship between the probability of a correct response to an item and the values of $\underset{\sim}{\theta}$ are given by the following function:

$$P(x = 1|\underset{\sim}{\theta}, \underset{\sim}{\sigma}) = \frac{e^{(\sigma_1\theta_1 + \sigma_2\theta_2 + \sigma_3)}}{1 + e^{(\sigma_1\theta_1 + \sigma_2\theta_2 + \sigma_3)}} , \qquad [3]$$

where σ_1, σ_2, and σ_3 are values related to the shape of the probability surface for this particular item. An example of the probability surface is given for $\sigma_1 = 1.5$, $\sigma_2 = .5$, and $\sigma_3 = .65$ in Figure 1. The difficulty function for this item is defined as

$$\frac{\delta^2 P(x = 1|\underset{\sim}{\theta}, \underset{\sim}{\sigma})}{\delta\theta_1^2} = 0 . \qquad [4]$$

The second derivative is only taken with respect to θ_1 in this case because the points of inflection define the same function in both dimensions. If for simplicity, P is used in place of $P(x = 1 | \underset{\sim}{\theta}, \underset{\sim}{\sigma})$, the second derivative is equal to

$$\frac{\delta^2 P}{\delta\theta_1^2} = \sigma_1{}^2 P(1 - 3P + 2P^2) . \qquad [5]$$

If this expression is set equal to zero and solved for P, three solutions result--0, .5, and 1. Since 0 and 1 are degenerate cases where $\underset{\sim}{\theta} = \underline{+}\infty$, the difficulty function is defined as the intercept of the probability surface with the .5 plane (a plane parallel to the θ plane at $P = .5$)

The line of intersection of the .5 plane with the probability surface can be obtained by determining which values of θ result in a .5 probability of a correct response. Since the exponent in Equation 3 must be equal to zero to obtain a probability of .5, the appropriate values of $\underset{\sim}{\theta}$ are the solutions to the equation

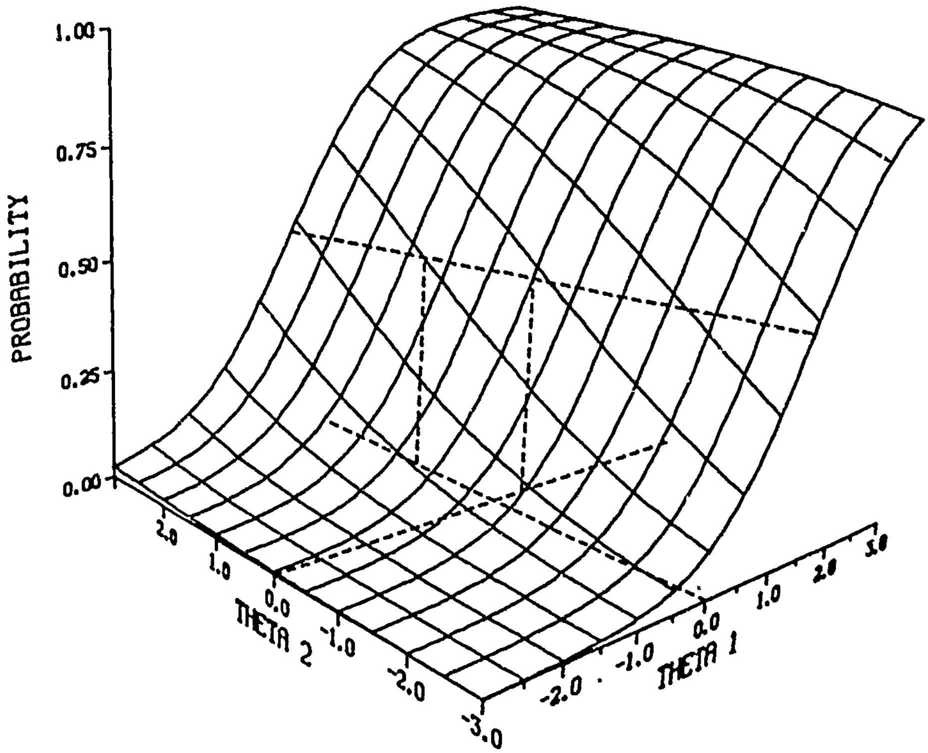
$$\sigma_1\theta_1 + \sigma_2\theta_2 + \sigma_3 = 0 . \qquad [6]$$

This is the equation of a straight line in the $\underset{\sim}{\theta}$ plane. In the usual linear form, the equation becomes

$$\theta_2 = -\frac{\sigma_1}{\sigma_2}\theta_1 - \frac{\sigma_3}{\sigma_2} . \qquad [7]$$

This line is shown as a dashed line on the .5 plane in Figure 1. Thus, for this example, the difficulty of the item is defined as a linear function instead of a single value.

Figure 1
An Example of a Two-Dimensional Probability Surface



The difficulty of the item in the usual latent trait sense can be determined on each dimension by holding the ability on the other dimension constant; for example, the point on the θ_2 scale that yields a .5 probability of a correct response when $\theta_1 = 0$ is $-\sigma_3/\sigma_2$, which is equal to $-(-.650/1.5) = .43$ for the surface given in Figure 1. Similarly, when $\theta_2 = 0$, $\theta_1 = -\sigma_3/\sigma_1 = -(-.650/.5) = 1.3$. Note that these "conditional difficulties" are different for each dimension even though there was only one summative term in the model, σ_3. When the dimensionality of the latent space is greater than two, the difficulty function for an item defines a hyperplane if a logistic model is used to describe the probability surface.

The definition of the discriminating power of the item in a multidimensional space can also be generalized from that used in the unidimensional case. In the unidimensional case, the discriminating power of an item is a function of the slope of the item response function at the point of inflection. The discriminating power of the item in a multidimensional space can likewise be determined by calculating the slope of the item response surface along the line defining the difficulty of the item.

This slope can be determined by evaluating the first derivative of the item response surface for values on the difficulty line. In the example given above, the first derivative of the item response surface with respect to θ_1 is

$$\frac{\partial P}{\partial \theta_1} = \sigma_1 P(1 - P) = \sigma_1 \left[\frac{e^{(\sigma_1\theta_1 + \sigma_2\theta_2 + \sigma_3)}}{1 + e^{(\sigma_1\theta_1 + \sigma_2\theta_2 + \sigma_3)}}\right]\left[\frac{1}{1 + e^{(\sigma_1\theta_1 + \sigma_2\theta_2 + \sigma_3)}}\right] . \quad [8]$$

If this function is evaluated for points on the line $\theta_2 = (-\sigma_1/\sigma_2)\theta_1 - \sigma_3$, the result is $.25\sigma_1$ for all values of θ_1. Thus, σ_1 is a discrimination parameter for the first dimension. Since the model is symmetric with respect to the θ's, the derivative with respect to θ_2 results in $.25\sigma_2$ when evaluated at all values on the difficulty line. Thus, both the difficulty function and the discrimination parameters are defined by the σ parameters in the exponent of this model. As with the difficulty function, the discrimination of a multidimensional item can be generalized to many dimensions. In the general case, the discrimination with respect to a dimension is the slope of the item response surface at the difficulty hyperplane. This slope may be a function of the $\underset{\sim}{\theta}$ vector in more complex models.

Up to this point, the complete latent space has been defined and the concept of an item response surface (IRS) has been introduced. Extensions of the unidimensional concepts of difficulty and discrimination have also been defined for the multidimensional IRS. The goal of this research, however, was to estimate the amount of ability an examinee possesses on each of these dimensions in the multidimensional latent space. Before this goal can be attained, two steps must be completed. First, a reasonable and convenient form for the IRS must be selected; and secondly, the parameters of the item response surface for each item must be determined.

Multidimensional Latent Trait Models

A number of models already exist in the literature for approximating the IRS. Each of these models will now be described, and the characteristics of the surface that they define will be summarized. Subsequently, one model will be selected for further analysis, leading to the estimation of model parameters.

Rasch's Model

The first of the models to be produced for approximating the IRS in a multidimensional space was presented by Rasch (1961). Although the model was not specifically designed to represent multidimensional data, Rasch indicated that vectors could be used for item and person parameters, thus extending the model to the multidimensional case.

The general form of the model is given by

$$P(x = i|\underset{\sim}{\sigma},\underset{\sim}{\theta}) = \frac{e^{\left(\sum_{\ell}^{n} w_{i\ell}\theta_\ell + \sum_{\ell}^{n} u_{i\ell}\sigma_\ell + \sum_{\ell}^{n}\sum_{m}^{n} v_{i\ell m}\theta_\ell\sigma_m + z_i\right)}}{\sum_{i}^{k} e^{\left(\sum_{\ell}^{n} w_{i\ell}\theta_\ell + \sum_{\ell}^{n} u_{i\ell}\sigma_\ell + \sum_{\ell}^{n}\sum_{m}^{n} v_{i\ell m}\theta_\ell\sigma_m + z_i\right)}} , \quad [9]$$

where

$\underline{x}$ is one of $i = 1, \ldots, k$ item responses;
$\underline{\theta}$ is the person parameter vector with elements θ_ℓ, $\ell = 1, \ldots, n$;
$\underline{\sigma}$ is the item parameter vector with elements σ_ℓ, $\ell = 1, \ldots, n$;
$\underline{w}$, $\underline{u}$, and $\underline{v}$ are weights for the person and item dimensions;
$\underline{z}$ is a scaling constant for the item responses; and
$\underline{n}$ is the number of dimensions.

The model can be shown more conveniently in vector form as

$$P(x = i|\underset{\sim}{\sigma},\underset{\sim}{\theta}) = \frac{e^{(\underset{\sim}{W}_i{}'\underset{\sim}{\theta} + \underset{\sim}{U}_i{}'\underset{\sim}{\sigma} + \underset{\sim}{\theta}'\underset{\sim}{V}_i\underset{\sim}{\sigma} + z_i)}}{\sum_i^k e^{(\underset{\sim}{W}_i{}'\underset{\sim}{\theta} + \underset{\sim}{U}_i{}'\underset{\sim}{\sigma} + \underset{\sim}{\theta}'\underset{\sim}{V}_i\underset{\sim}{\sigma} + z_i)}}, \qquad [10]$$

where $\underset{\sim}{W}$ and $\underset{\sim}{U}$ are vectors of weights for each item response, and $\underset{\sim}{V}$ is a matrix of weights.

This model is extremely general, allowing both dichotomous and polychotomous scoring and containing both the 1-parameter and 2-parameter logistic models as special cases. Because the model is so general, it is difficult to determine the form of the item difficulty and discrimination functions. However, for the special case of a dichtomously scored item with $\underset{\sim}{W}$ and $\underset{\sim}{U}$ equal to unit vectors for a correct response and zero vectors for an incorrect response, $\underset{\sim}{V}$ equal to the identity matrix for a correct response and a zero matrix for an incorrect response, and $\underline{z}$ equal to zero for all responses, the difficulty function is a hyperplane and the conditional slopes of the surface where it intersects the .5 plane are functions of σ_i. The model presented in Equation 3 and shown in Figure 1 is a special case of the general model.

Only one study is known that uses this general model to represent multidimensional item response data (Reckase, 1972), although there have been other applications of the model (Andersen, 1982; Andrich, 1978). In the Reckase (1972) study an attempt was made to estimate the parameters using a least squares procedure for a special case of this model, where $\underset{\sim}{V}_i$ is a zero matrix for all responses and z_i takes on zero values for all responses. The attempt was not entirely successful, however, in that the fit of the multivariate model to multivariate data was no better than the fit of the simple Rasch (1960) model to the same data when estimates of the parameters were used. The poor results were attributed to two factors: (1) both the parameter vectors and the weights were estimated from the data and (2) the sample size used was too small to estimate accurately the large number of unknown variables. The parameter estimates were interpretable, however, suggesting that a less ambitious approach might be fruitful.

Mulaik's Model

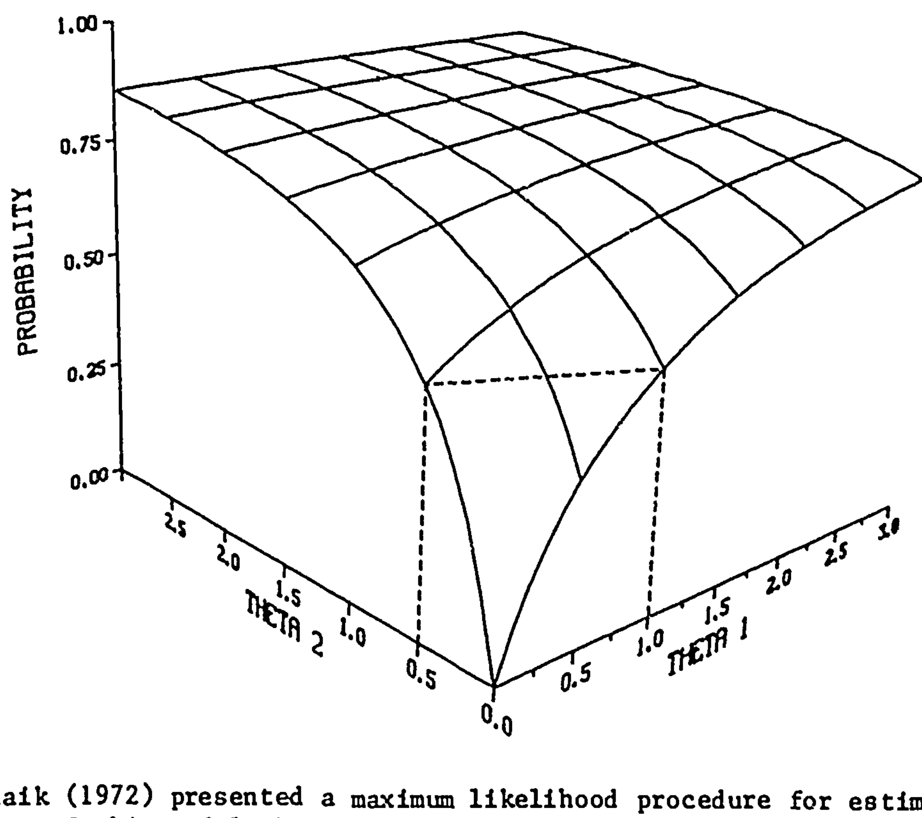
Another multidimensional model that was developed as an extension of the work of Rasch (1960) was proposed by Mulaik (1972). This model is given by

$$P(x = 1|\underset{\sim}{\gamma},\underset{\sim}{\pi}) = \frac{\sum_{i=1}^{n} \gamma_i \pi_i}{1 + \sum_{i=1}^{n} \gamma_i \pi_i} , \qquad [11]$$

where the γ_i's are item parameters and the π_i's are person parameters for the interaction of a person and an item in an *n*-dimensional space.

The previous definitions of the item difficulty and discrimination do not apply to this model, since the surface defined by Equation 11 does not have a point or line of inflection. However, the intersection of the surface with the .5 plane is a hyperplane and could be used to define item difficulty. Unlike the previous model, the conditional slope of the IRS at the intersection with the .5 plane is not simply a function of the item parameters but also depends on the ability parameters. A two-dimensional example of the response surface described by this model is presented in Figure 2.

Figure 2
Item Response Surface for Mulaik's Model



Mulaik (1972) presented a maximum likelihood procedure for estimating the parameters of this model, but it appears that it has not ever been applied. He cautioned that the amount of computation and the constraints required to estimate the parameters may be too great for the current generation of computers.

Sympson's Model

A third model that has been developed to describe the interaction of a person and an item in a multidimensional latent space was described by Sympson (1978). Rather than extend the 1-parameter logistic model, as done by Rasch (1961) and Mulaik (1972), Sympson (1978) based his model on an extension of the 3-parameter logistic model (Birnbaum, 1968). The mathematical expression for this model is given by

$$P(x = 1|\underset{\sim}{\theta},\underset{\sim}{a},\underset{\sim}{b},c) = c + (1 - c) \prod_{\ell=1}^{n} \left[1 + e^{[-1.7a_\ell(\theta_\ell - b_\ell)]}\right]^{-1}, \qquad [12]$$

where

x is the item response,
$\underset{\sim}{\theta}$ is a vector of ability parameters,
$\underset{\sim}{a}$ is a vector of discrimination parameters,
$\underset{\sim}{b}$ is a vector of difficulty parameters, and
c is a pseudo-chance level parameter.

An example of the surface defined by Equation 12 is given in Figure 3 for the two-dimensional case with parameters $c = .2$, $a_1 = .7$, $a_2 = 1.2$, $b_1 = -.6$, and $b_2 = .5$.

Unlike the models presented by Rasch (1961) and Mulaik (1972), the root of the second derivative of this equation does not define a difficulty function but gives a single value for each dimension. This value is simply the b parameter for that dimension. The difficulty of an item using this model can therefore be defined as the vector of b values, which defines a point in the multidimensional space.

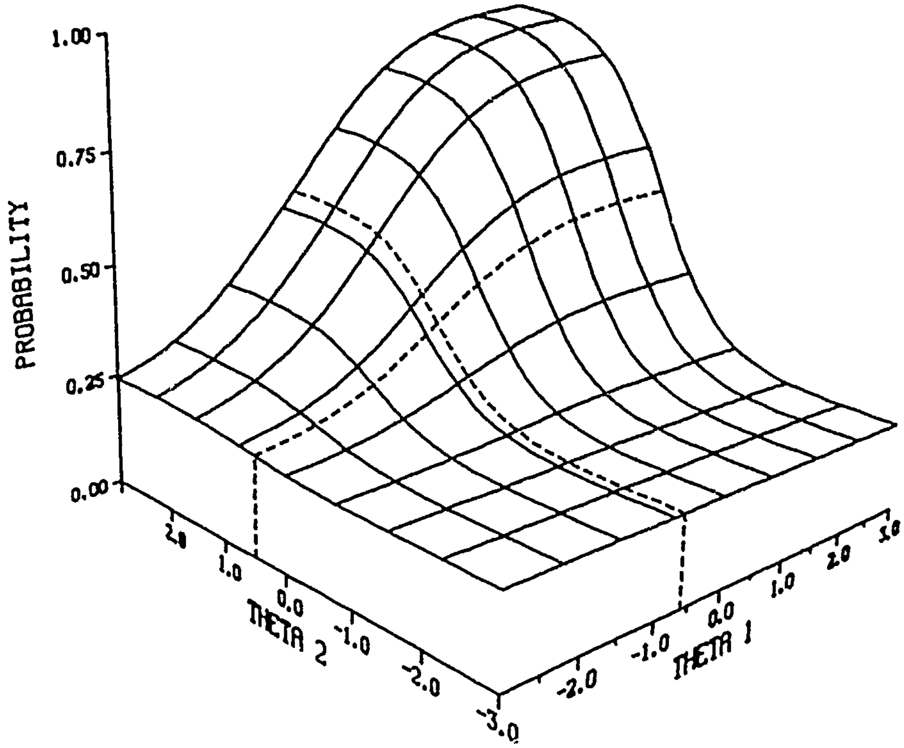
The slope of the IRS at the point of inflection for Sympson's (1978) model is given by

$$\frac{(1-c)(1.7)^n \prod_{\ell=1}^{n} a_\ell}{4^n},$$

which is solely a function of the item discrimination parameters and the pseudo-chance level parameter. If the slope of the function is determined at the difficulty point with respect to just one dimension, θ_i, the result is $(1 - c)(1.7)a_i/4 \times 2^{(n-1)}$, where n is the number of dimensions. Thus for this model the $\underset{\sim}{a}$ vector defines the discrimination power of the item.

Sympson (1978) has done some preliminary work on estimating the parameters of this model for some simple cases, but no procedure has yet been published for the full multidimensional case. Lord (1978), in discussing Sympson's (1978)

Figure 3
Item Response Surface for Sympson's Model



paper, has suggested that a Bayesian or maximum likelihood procedure might be more fruitful than the method Sympson proposed. However, these methods have not been developed to the point where this model can be applied to actual test data.

Bock and Aitkin's Model

Bock and Aitkin (1981) suggested a multidimensional latent trait model that is an extension of the 2-parameter normal ogive model (Lord & Novick, 1968). They also indicated that this model is similar to the factor analysis procedures developed for dichotomous data by Christoffersson (1975) and Muthén (1978).

The mathematical form of the Bock and Aitkin (1981) model is given by the equation

$$P(x = 1|\underset{\sim}{\theta},\underset{\sim}{a},c) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{z(\underset{\sim}{\theta})} e^{\left(-\frac{t^2}{2}\right)} dt , \qquad [13]$$

where

$z(\underset{\sim}{\theta}) = c + \Sigma_{i=1}^{n} a_i \theta_i$,

$\underset{\sim}{\theta}$ is a vector of ability parameters,

$\underset{\sim}{a}$ is a vector of discrimination parameters, and

$\underline{c}$ is the item difficulty parameter.

Due to the similarity of the cumulative model and logistic functions, this model is very similar to one of the special cases of the general Rasch model given in Equation 9. Specifically, it corresponds to the case where the exponent of that model is equal to $\sum_i \sigma_i \theta_i + \sigma_{n+1}$ for a correct response and 0 for an incorrect response.

As with the general Rasch model, the difficulty function for the model is defined by the equation for a hyperplane

$$\Sigma a_i \theta_i + c = 0 \ . \qquad [14]$$

The conditional slope of the IRS for points on the difficulty function is given by $a_i/\sqrt{2\pi}$, demonstrating that the $\underline{a}$ vector is related to the discriminating power of the item. Since the two-dimensional IRS for this model is indistinguishable from the logistic surface presented in Figure 1, an example is not presented for this model.

Bock and Aitkin (1981) have produced an estimation procedure for their model based on the marginal maximum likelihood technique. They have applied this procedure to data assuming a two-dimensional solution. The results of the analysis showed that different quadrature procedures used in conjunction with marginal maximum likelihood techniques gave slightly different results but that a two-dimensional solution seemed to fit the data fairly well. No other applications of this model are known.

Samejima's Model

Samejima (1974) presented a more general version of the model suggested by Bock and Aitkin (1981) as a special case of her continuous response model in a multidimensional latent space. Her model is given by the equation

$$P^*_{z_g}(\underset{\sim}{\theta}) = \int_{-\infty}^{\underset{\sim}{a}_g'(\underset{\sim}{\theta} - \underset{\sim}{b}_g)} \psi_g(u)du \ , \qquad [15]$$

where

z_g is the point of dichotomy of the continuous trait measured by this item,

$P^*_{z_g}(\underset{\sim}{\theta})$ is the probability of a correct response,

$\underset{\sim}{a}_g$ is a vector of discrimination parameters,

$\underset{\sim}{\theta}_g$ is a vector of ability parameters,

$\underset{\sim}{b}_g$ is a vector of difficulty parameters, and

$\psi_g(u)$ is a twice differentiable function.

When ψ_g is the normal density function, Equation 15 is identical to Equation 13 with $c = -\Sigma a_i b_i$. When ψ_g is defined as the logistic density function, the model is a special case of the general Rasch model given in Equation 9. Samejima also points out the similarity of the model to linear factor analysis.

Whitely's Models

One other class of multidimensional latent trait models exists in the literature, but this class of models was developed from a different perspective than the others. The models presented up to this point generally consider the dimensions required to determine the complete latent space as unknown hypothetical constructs, the properties of which need to be discovered. In contrast, the class of the models proposed by Whitely (1980) considers the dimensions to be components in a cognitive model of performance. These dimensions are defined in advance of being estimated as particular cognitive processes.

The mathematical form of the model used by Whitely is similar to that used by Sympson (1981) in that it is composed of the product of separate logistic model terms. The particular model, called the multicomponent latent trait model, is given by the equation

$$P(x = 1|\underset{\sim}{\theta},\underset{\sim}{b}) = \prod_{i=1}^{n} \frac{e^{(\theta_i - b_i)}}{1 + e^{(\theta_i - b_i)}} , \qquad [16]$$

where the variables are as defined above. This model is the same as the Sympson (1978) model with $c = 0$ and $a_i = 1/1.7$ for all $\underline{i}$. Since it is a special case of the Sympson model, it also has a difficulty function equal to the $\underline{b}$ vector, and the slope at the point defined by the $\underline{b}$ vector is $1/4^n$. Whitely (1980) has also produced more complex versions of this model, but they are all composed of combinations of the 1-parameter logistic model.

Whitely (1980) has developed procedures for estimating the parameters of this model, but the estimation has been performed in a different manner than the other models that have been described. Whereas the estimation procedures for the other models have attempted to estimate the vector parameters from the dichotomous responses to test items, Whitely has developed an experimental design for collecting responses on each cognitive component separately. The parameters of each of the product terms of the model are then estimated separately using procedures developed for the unidimensional Rasch model. No descriptions of a procedure for simultaneous estimation of all of the parameters of the model has been found in the literature.

Comparison of the Multidimensional Models

An analysis of the six models that have been described above indicated that these models fall into three basic classes. The first of these classes (Class I) is of the form

$$P(x = 1|\underset{\sim}{\theta},\underset{\sim}{\sigma}) = \int_{-\infty}^{z(\underset{\sim}{\theta})} \psi(u)\,du , \qquad [17]$$

where $z(\theta)$ is a linear function of the elements of $\underset{\sim}{\theta}$. The general Rasch (1961) model, the Bock and Aitkin (1981) model, and the Samejima (1974) model fall into this class. All of these models allow high ability on one dimension to compensate for low ability on another dimension, resulting in what Sympson (1978) has labeled as compensatory models. All of these models have linear difficulty functions when used with dichotomously scored data and have conditional slopes at points on the difficulty function that are functions of the corresponding discrimination parameters. These models are fairly simple from a mathematical point of view (especially when $\psi(u)$ is the logistic density function), and estimation procedures have been developed for the parameters (Bock & Aitkin, 1981).

Class II models contain a single example, the model proposed by Mulaik (1972). This model is of the form

$$P(x = 1|\underset{\sim}{\gamma},\underset{\sim}{\pi}) = \frac{\sum_{i=1}^{n} \gamma_i \pi_i}{1 + \sum_{i=1}^{n} \gamma_i \pi_i}, \qquad [18]$$

where the variables are as defined earlier. This model is also compensatory in Sympson's (1978) sense, but it is unlike the previous class in that the ability metric is defined from 0 to $+\infty$ instead of from $-\infty$ to $+\infty$. This results in an IRS that does not have a point or line of inflection. If the difficulty of the item is defined by the intersection of the item with the .5 plane, the result is a linear function similar to that for the Class I models. The slope of the IRS at the difficulty function has the property of changing with its position relative to the ability dimensions. Mulaik (1972) has proposed an estimation procedure for this model, but there have been no studies to determine its practicality.

Class III models contain the models proposed by Sympson (1978) and Whitely (1980). These models take the form

$$P(x = 1|\underset{\sim}{\theta},\underset{\sim}{\sigma}) = \sigma_1 + (1 - \sigma_1) \prod_{i=1}^{n} P'_i(\theta_i), \qquad [19]$$

where σ_1 is a lower asymptote parameter and $P'_i(\theta_i)$ is the probability of response with respect to a specific dimension. In Whitely's (1980) model the dimensions are defined as specific cognitive processes required to solve the problem proposed in the item, and the σ_1 parameter is assumed to be zero. In Sympson's (1978) model the dimensions are hypothetical traits based on commonalities among items.

Unlike the previous models, the Class III noncompensatory models do not allow a high ability on one dimension to compensate for a lower ability on another. The lowest of the values of $P'_i(\theta_i)$ defines the upper bound of $P(x = 1|\underset{\sim}{\theta},\underset{\sim}{\sigma})$. Although some work has been done on the estimation of parameters for the Class III models, no generally accepted algorithm for estimation of the parameters of these models is known to exist.

Several issues need to be considered in selecting one of these models as a description of the interaction between a person and an item. The first is whether the model is realistic. This depends on whether a compensatory or non-compensatory model is appropriate for actual persons and items. Unfortunately, this is a question that still needs to be answered, based on research in cognitive psychology. Although sufficient information is presently not available, the applicability of the models to actual testing situations may provide an answer.

Beyond questions of the psychological meaningfulness of the models are questions of practicality. The most well developed estimation procedures are available for the Class I models; and these models tend to have the most flexible options due to the characteristics of the exponential term. As a consequence, the Class I models tend to be more promising than the other models. Of the Class I models, the generalized Rasch model has the greatest flexibility in its options and is the most mathematically tractable. The remainder of this paper will therefore concentrate on the properties of this model and the procedures for the estimation of its parameters.

Application of the General Rasch Model

Although the general Rasch model is a generalization of the 1-parameter logistic model, a very simple model, in its most general form the model is very complex. A study (McKinley & Reckase, 1982) was thus undertaken to determine whether a less complex formulation of the model would be adequate for modeling multidimensional response data.

Method

Design. The general design of this study was to first evaluate the properties of a simple formulation of the general model and then to evaluate increasingly more complex versions of the model produced by inserting additional terms. The initial form of the model investigated is given by Equation 20:

$$P(x|\underset{\sim}{\theta}_j,\underset{\sim}{\sigma}_i) = \frac{1}{\gamma(\underset{\sim}{\theta}_j,\underset{\sim}{\sigma}_i)} \exp(\underset{\sim}{U}'\underset{\sim}{\sigma}_i + \underset{\sim}{W}'\underset{\sim}{\theta}_j) \ . \qquad [20]$$

For each level of model complexity, the properties of the model were investigated and the reasonableness and usefulness of the model were explored. This was done primarily by generating simulated test data to fit the particular form of the model being investigated and by analyzing that data in an attempt to assess how well the characteristics of the data matched the characteristics of real test data. If it were found that a particular form of the model could not be used to generate realistic data in terms of either dimensionality or item characteristics, then that form of the model was rejected and a different form of the model was investigated. Distinct special cases of the model were obtained by eliminating different terms from the general model by setting the appropriate parameter weights equal to zero.

Analyses. The analyses of the generated data that were performed included factor analysis and traditional item analysis. The purposes of the analyses were three-fold. One purpose was to determine whether the obtained factor structure of the data resembled the factor structure typically obtained for real test data. The second purpose was to determine whether the obtained unidimensional item characteristics (difficulty and discrimination) were similar to those obtained for real data. The third purpose was to aid in the interpretation of the parameters of the model.

If it were found that a model could not be used to generate realistic data, an attempt was made to determine what changes in the model would yield a more acceptable model. In many cases it was necessary to generate additional data, using different values for the parameters of the model in order to answer specific questions about a particular model statement. Once an understanding was gained as to the roles played by different parameters of the model, predictions could be made regarding the effects of adding or eliminating other terms.

Results. As a result of the analyses performed on the different formulations of the model, a good understanding of the significance of the terms in the model was gained. It is now clear that parameters play quite varied roles depending on the term of the model in which they appear. Because of this, the characteristics of the data for which the model can be used vary markedly, depending on the form of the model.

To begin with, it is clear that the use of $\underset{\sim}{W}'\underset{\sim}{\theta}_j$ and $\underset{\sim}{U}'\underset{\sim}{\sigma}_i$ terms alone is not sufficient for modeling multidimensional response data. The linear composite represented by the $\underset{\sim}{U}'\underset{\sim}{\sigma}_i$ term in the model determines only item difficulty. Moreover, the order of the σ_i vector is unimportant. It is the magnitude of the inner product of the item parameter vector and the weight vector that determines the difficulty of the item. Regardless of whether the vectors have one or five elements, as long as the inner product is the same, the difficulty of the item in terms of proportion of correct responses is the same.

It is also clear from the results of the analyses that the product term $\underset{\sim}{\theta}_j'\underset{\sim}{V}\underset{\sim}{\sigma}_i$ is necessary if item discrimination is to be modeled. When data are generated using only the inner product terms, the items modeled have constant discriminations and the resulting data are unidimensional. When the product term is included, the items modeled have varying discrimination. Moreover, the factor analysis results indicate that the dimensionality of the generated data is determined by the number of elements from the σ_i vector used in the $\underset{\sim}{\theta}_i'\underset{\sim}{V}\underset{\sim}{\sigma}_i$ term. However, it should be emphasized that if the $\underset{\sim}{V}$ matrix contains more than one nonzero element in a row or column, a σ or θ term will appear in the exponent multiplied by more than one of the θ or σ parameters, respectively (e.g., $\theta_1\sigma_1 + \theta_1\sigma_2$ or $\theta_1\sigma_1 + \theta_2\sigma_1$). The presence of these terms in the exponent results in difficulty in determining the meaning of the θ and σ vectors.

The elements in the σ_i vector in the $\underset{\sim}{\theta}_i'\underset{\sim}{V}\underset{\sim}{\sigma}_i$ term determine the discrimination of the modeled items. Because of this, use of the same elements of the σ_i

vector in both the $\theta_i'V\sigma_i$ term and the $U'\sigma_i$ term produces an undesirable characteristic in the data. Since the elements used in the $\theta_i'V\sigma_i$ term determine item discrimination, while the elements in the $U'\sigma_i$ term determine item difficulty, use of the same elements in both terms yields items having highly related observed item difficulty values and item discrimination values. This is not a very realistic situation.

Conclusions. On the basis of the results of the analyses, it was concluded that if the model is to be used to represent multidimensional item response data, it must include the $\theta_i'V\sigma_i$ term, but no element in either the σ_i or the θ_j vector should be multiplied by more than one term in the other vector. If items are to vary in difficulty, the $U'\sigma_i$ term must be included; but to avoid highly related values for unidimensional measures of item discrimination and difficulty, no element of the σ_i vector should appear in both the $U'\sigma_i$ term and the $\theta_i'V\sigma_i$ term. The model that appeared to be most useful for modeling multidimensional response data is given by

$$P(x|\theta_j, \sigma_i) = \frac{1}{\gamma(\theta_j, \sigma_i)} \exp(U'\sigma_i + \theta_j'V\sigma_i) , \quad [21]$$

where no elements of the σ_i vector appear in both terms of the model.

There was one additional significant finding. Although it was concluded that the use of the model without the $\theta_j'V\sigma_i$ term was unsuccessful in modeling multidimensional response data, this result was obtained when the special case of the model was applied to dichotomously scored item response data. This model may also be applied to polychotomously scored item response data. In one specific application of this model to polychotomous response data, some measure of success was attained in modeling multidimensional data using only the $U'\sigma_i$ and $W'\theta_j$ terms. Dichotomously scored items were transformed to polychotomous form by grouping items together to form clusters having several nominal response categories. When these data were analyzed, several dimensions could be determined. However, this approach has not been extensively investigated, and any conclusions drawn as to the usefulness of this approach are at best tentative.

Estimation of Parameters in the General Rasch Model

Two basic approaches to estimating item parameters can be distinguished. One approach is to specify a distribution of the latent ability of the population from which the sample was taken and then to integrate the response function with respect to that distribution to obtain item parameters unconditionally (Bock, 1972). This approach has been taken by Bock and Lieberman (1970) and Bock and Aitkin (1981). The other approach is to estimate item parameters by treating the examinees' abilities as fixed unknowns and by conditioning item parameter estimation on estimates of ability (Bock, 1972). This approach has been taken by Lord (1968) and Kolakowski and Bock (1970). The present research considers both approaches. However, at this time only the conditional item parameter estimation procedure is designed for the most general form of the model; it is still limited to the use of dichotomously scored data and it does not estimate the $\underset{\sim}{W}$, $\underset{\sim}{U}$, and $\underset{\sim}{V}$ parameter weights. The unconditional item parameter estimation procedure is designed for use with the form of the model given by Equation 21. As was the case with the conditional item parameter estimation procedure, this procedure does not estimate the parameter weights.

In addition to the two procedures that have been developed for item parameter estimation, a maximum likelihood ability estimation procedure has been developed for estimating the ability parameters for the general Rasch model. This procedure estimates ability conditionally and is combined with the conditional maximum likelihood item parameter estimation procedure discussed above to form a conditional maximum likelihood estimation procedure for simultaneous estimation of the item and ability parameters of the general Rasch model. Used alone, the conditional maximum likelihood ability estimation procedure can be used to estimate ability using the item parameter estimates obtained from the unconditional item parameter estimation procedure.

Unconditional Item Parameter Estimation

General procedure. The unconditional item parameter estimation procedure is an adaptation of a procedure proposed by Bock and Aitkin (1981), which was designed for use with a multidimensional 2-parameter normal ogive model (see Equation 13). In the initial step of this procedure, a distribution of ability is assumed, and quadrature nodes and weights are selected for use in determining expected sample sizes for portions of the distribution using numerical integration. For the multidimensional case, the prior distribution of ability is multivariate, and the nodes and weights are vectors. At each node the expected number of examinees from the sample having the ability represented by the node is computed, as is the expected frequency of correct responses to each item by examinees with the ability represented by the node. These expected number-correct scores and expected sample sizes are used in a logit analysis, which is performed using a least squares regression procedure. The results of the logit analysis are estimates of the parameters of the model.

The initial stage of the estimation procedure requires provisional estimates of the item parameters. These provisional estimates are used in the first step of the initial stage, which involves obtaining expected sample sizes and number-correct scores. In the second step of the initial stage, a logit analysis is performed to obtain new estimates of the item parameters. These new estimates are used in the first step of the second stage, which involves obtaining new estimates of sample sizes and number-correct scores. These new sample sizes and number-correct scores are used in another logit analysis, which yields a new set of item parameter estimates. These stages are repeated until a criterion of convergence is met or until a limit on the number of stages is reached.

Expected sample sizes and number-correct scores. The expected sample size at each node is given by

$$\bar{N}_k = \sum_{\ell=1}^{N} \frac{L_\ell(\underset{\sim}{\theta}_k)\underset{\sim}{W}_k}{\tilde{P}_\ell}, \qquad [22]$$

where

$\bar{N}_k$ is the expected sample size at node $\underline{k}$;

$\underset{\sim}{\theta}_k$ is the ability represented by node $\underline{k}$;

$\underset{\sim}{W}_k$ is the weight for node $\underline{k}$;

N is the number of examinees in the sample; and

$L_\ell(\underset{\sim}{\theta}_k)$ is the likelihood of response vector ℓ given node $\underline{k}$.

$L_\ell(\underset{\sim}{\theta}_k)$ is given by

$$L_\ell(\underset{\sim}{\theta}_k) = \prod_{i=1}^{n} P(u_{ij}) \ , \qquad [23]$$

where $P(u_{i\ell})$ is given by

$$P(u_{il}) = \exp[u_{i\ell}(c_i + \underset{\sim}{A}_i\underset{\sim}{\theta}_k)] \ / \ (1 + \exp[c_i + \underset{\sim}{A}_i\underset{\sim}{\theta}_k]) \ , \qquad [24]$$

where

$u_{i\ell}$ is the response to item $\underline{i}$ in response vector ℓ,

c_i is the difficulty parameter for item $\underline{i}$, and

$\underset{\sim}{A}_i\underset{\sim}{\theta}_k$ is given by

$$\underset{\sim}{A}_i\underset{\sim}{\theta}_k = \sum_{m=1}^{M} a_{im}\theta_{km} \ , \qquad [25]$$

where

M is the number of dimensions,

a_{im} is the $\underline{m}$th element of the discrimination parameter vector for item $\underline{i}$, and

θ_{km} is the $\underline{m}$th element of node $\underline{k}$.

$\tilde{P}_\ell$ is given by

$$\tilde{P}_\ell = \sum_{k=1}^{q} L_\ell(\underset{\sim}{\theta}_k)\underset{\sim}{W}_k \ , \qquad [26]$$

where $\underline{q}$ is the number of quadrature nodes. The sum over all of the nodes of the ratio in Equation 22 is one, and the sum of the $\bar{N}_k$ over all of the nodes is N, where N is the number of examinees in the sample.

The expected number-correct score for item $\underline{i}$ at node $\underline{k}$, r_{ik}, is given by

$$\bar{r}_{ik} = \sum_{\ell=1}^{s} \frac{u_{i\ell}L_\ell(\underset{\sim}{\theta}_k)\underset{\sim}{A}_k)}{\tilde{P}_\ell} \ , \qquad [27]$$

where the other terms are as previously defined. The sum over all of the nodes of the r_{ik} for an item is equal to the observed number-correct score for the item.

Multiple logit analysis. For each item the expected proportion-correct score at each node is given by

$$\hat{P}_{ik} = \bar{r}_{ik} / \bar{N}_k , \qquad [28]$$

where all of the terms are as previously defined. The $\hat{P}_{ik}$ are converted to logits and used as the dependent variable in a regression analysis. The independent variables in the regression analysis are the elements in the node vectors. The model for the regression analysis is given by

$$\log_e[\hat{P}_{ik} / (1 - \hat{P}_{ik})] = c_i + \underset{\sim}{A}_i \underset{\sim}{\theta}_k + \text{error} , \qquad [29]$$

where all of the terms are as previously defined. The regression analysis results in estimates of c_i and $\underset{\sim}{A}_i$.

Conditional Item Parameter Estimation

This procedure is based on the conditional maximum likelihood estimation technique. The procedure begins with the computation of an initial weighted item score on each of M dimensions using

$$X_{ik} = \sum_{j=1}^{N} u_{jk} + \sum_{j=1}^{N} \theta_{jk} v_{jk} , \qquad [30]$$

where

X_{ik} is the initial score for item i on dimension k,

u_{jk} is the kth element of $\underset{\sim}{U}$, and

v_{jk} is the kth element of $\underset{\sim}{V}$.

The index j indicates that the value of the elements of $\underset{\sim}{U}$ and $\underset{\sim}{V}$ are dependent on the response of the jth examinee to item i. These scores are converted to z scores via the transformation

$$z_{ik} = (X_{ik} - \bar{X}_k) / s_k , \qquad [31]$$

where

z_{ik} is the z score for item i on dimension k,

X_{ik} is as defined in Equation 30,

$\bar{X}_k$ is the mean of the weighted scores on dimension k, and

s_k is the standard deviation of the weighted scores on dimension k.

The z scores are used as initial estimates of the item parameters. That is,

$$\hat{\sigma}_{oik} = z_{ik} , \qquad [32]$$

where $\hat{\sigma}_{oik}$ is the item parameter estimate of item i on dimension k after 0 iterations.

New item parameter estimates are obtained using the initial item parameter estimates as the starting point in an iterative process. One iteration is complete when a new item parameter estimate is obtained for each item on each dimension. Within a single iteration, new estimates on the first dimension are obtained while holding the estimates on all other dimensions constant at the values obtained on the previous iteration. Estimates on the second dimension are obtained using previous estimates on all other dimensions except the first dimension. For the first dimension the new estimates are used. An iteration is complete when new estimates have been obtained on all dimensions. Iterations continue until a criterion of convergence is met or until a limit on the number of iterations is reached.

On the ℓth iteration, the new estimate on the <u>k</u>th dimension for item <u>i</u> is given by

$$\hat{\sigma}_{\ell ik} = \hat{\sigma}_{(\ell-1)ik} + \left[\frac{\partial}{\partial\sigma_{ik}} \log_e L(\sigma_{ik})\right] \Big/ \left[-\frac{\partial^2}{\partial\sigma_{ik}^2} \log_e L(\sigma_{ik})\right], \qquad [33]$$

where $\hat{\sigma}_{(\ell-1)ik}$ is the estimate for item <u>i</u> on dimension <u>k</u> from the previous iteration and log L is the log to the base <u>e</u> of the likelihood function for the response vector for item <u>i</u>. The likelihood function is given by

$$L(\sigma_{ik}) = \prod_{j=1}^{N} P(u_{ij}), \qquad [34]$$

where $P(u_{ij})$ is the probability of response u_{ij} by person <u>j</u> to item <u>i</u> with parameter σ_{ik}. $P(u_{ij})$ is given by

$$P(u_{ij}=1) = \frac{\exp(\underset{\sim}{W}_1{}'\underset{\sim}{\theta}_j + \underset{\sim}{U}_1{}'\underset{\sim}{\sigma}_i + \underset{\sim}{\theta}_j{}'\underset{\sim}{V}_1\underset{\sim}{\sigma}_i + z_1)}{\exp[\underset{\sim}{W}_o{}'\underset{\sim}{\theta}_j+\underset{\sim}{U}_o{}'\underset{\sim}{\sigma}_i+\underset{\sim}{\theta}_j{}'\underset{\sim}{V}_o\underset{\sim}{\sigma}_i+z_o] + \exp[\underset{\sim}{W}_1{}'\underset{\sim}{\theta}_j+\underset{\sim}{U}_1{}'\underset{\sim}{\sigma}_i+\underset{\sim}{\theta}_j{}'\underset{\sim}{V}_1\underset{\sim}{\sigma}_i+z_1]}, \qquad [35]$$

where all of the terms are as previously defined. The 0 and 1 subscripts on the vectors of weights, $\underset{\sim}{U}$, $\underset{\sim}{V}$, and $\underset{\sim}{W}$, indicate the values taken by those vectors for an incorrect and a correct response, respectively.

The first derivative of the $\log_e$ likelihood function is given by

$$\frac{\partial}{\partial\sigma_i} \log_e L(\underset{\sim}{\sigma}_i) = \sum_{j=1}^{N} \underset{\sim}{U}_j + \sum_{j=1}^{N} \underset{\sim}{\theta}_j{}'\underset{\sim}{V}_j - \sum_{j=1}^{N} (\underset{\sim}{U}_o+\underset{\sim}{\theta}_j{}'\underset{\sim}{V}_o)Q_{ij} - \sum_{j=1}^{N} (\underset{\sim}{U}_1+\underset{\sim}{\theta}_j{}'\underset{\sim}{V}_1)P_{ij}, \qquad [36]$$

where all of the items are as previously defined. The second derivative of the $\log_e$ likelihood function is given by

$$\frac{\partial^2}{\partial\sigma_i^2} \log_e L(\underset{\sim}{\sigma}_i) = \sum_{j=1}^{N} [\underset{\sim}{U}_o + \underset{\sim}{U}_1 + \underset{\sim}{\theta}_j{}'\underset{\sim}{V}_o + \underset{\sim}{V}_o + \underset{\sim}{V}_1]^2 P_{ij}Q_{ij}, \qquad [37]$$

where P_{ij} is the probability of a correct response to item $\underline{i}$ by person $\underline{j}$, $Q_{ij} = 1 - P_{ij}$, and the other terms are as previously defined.

Conditional Ability Parameter Estimation Procedure

The conditional ability estimation procedure is also a maximum likelihood estimation procedure. It is very similar to the conditional item parameter estimation procedure. For each examinee an initial weighted score is computed on each of M dimensions as

$$X_{jk} = \sum_{i=1}^{n} w_{jk} + \sum_{i=1}^{n} v_{jk}\sigma_{ik} , \qquad [38]$$

where X_{jk} is the initial score for person $\underline{j}$ on dimension $\underline{k}$. These scores are converted to $\underline{z}$ scores via the transformation

$$z_{jk} = (X_{jk} - \overline{X}_k) / s_k , \qquad [39]$$

where

z_{jk} is the $\underline{z}$ score for person $\underline{j}$ on dimension $\underline{k}$,

X_{jk} is as defined in Equation 38,

$\overline{X}_k$ is the mean of the weighted scores on dimension $\underline{k}$, and

s_k is the standard deviation of the weighted scores on dimension $\underline{k}$.

The $\underline{z}$ scores are used as initial estimates of ability. That is,

$$\hat{\theta}_{ojk} = z_{jk} , \qquad [40]$$

where $\hat{\theta}_{ojk}$ is the ability estimate of person $\underline{j}$ on dimension $\underline{k}$ after 0 iterations.

Estimates of ability are obtained using the same iterative process as was described for the conditional item parameter estimation procedure. On the ℓth iteration, the new estimate on the $\underline{k}$th dimension for person $\underline{j}$ is given by

$$\hat{\theta}_{\ell jk} = \hat{\theta}_{(\ell-1)jk} + \left[\frac{\partial}{\partial\theta_{jk}} \log_e L(\theta_{jk})\right] \Big/ \left[-\frac{\partial^2}{\partial\theta_{jk}^2} \log_e L(\theta_{jk})\right] , \qquad [41]$$

where $\theta_{(\ell-1)jk}$ is the estimate for person $\underline{j}$ on dimension $\underline{k}$ from the previous iteration and log L is the log to the base $\underline{e}$ of the likelihood function for the response vector for person $\underline{j}$. The likelihood function is given by

$$L(\theta_{ik}) = \prod_{i=1}^{n} P(u_{ij}) , \qquad [42]$$

where $P(u_{ij})$ is the probability of a response u_{ij} to item i by person j with ability θ_{jk}. $P(u_{ij})$ is given by Equation 35. In evaluating $\underline{\theta}_j$ the most recent item parameter estimates are used.

The first derivative of the log likelihood function is given by

$$\frac{\partial}{\partial \underset{\sim}{\theta}_j} \log_e L(\underset{\sim}{\theta}_j) = \sum_{i=1}^{n} \underset{\sim}{W}_i + \sum_{i=1}^{n} \underset{\sim}{V}_i \underset{\sim}{\sigma}_i - \sum_{i=1}^{n} (\underset{\sim}{W}_o + \underset{\sim}{V}_o \underset{\sim}{\sigma}_i) Q_{ij} - \sum_{i=1}^{n} (\underset{\sim}{W}_1 + \underset{\sim}{V}_1 \underset{\sim}{\sigma}_i) P_{ij} , \qquad [43]$$

where all the terms are as previously defined. The second derivative is given by

$$\frac{\partial^2}{\partial \underset{\sim}{\theta}_j^{\,2}} \log_e L(\underset{\sim}{\theta}_j) = \sum_{i=1}^{n} \left[\underset{\sim}{W}_o + \underset{\sim}{W}_1 + \left(\underset{\sim}{V}_o + \underset{\sim}{V}_1 \right) \underset{\sim}{\sigma}_i \right]^2 P_{ij} Q_{ij} , \qquad [44]$$

where, again, all the terms are as previously defined.

Evaluaton of the Estimation Procedures

General Design

The general approach taken to evaluate the estimation procedures was to apply the procedures to test data generated to fit a two-dimensional version of the model given by Equation 24 and then to compare the estimates of the parameters with the values used to generate the simulation data. For this purpose a data set comprising response data for 50 items and 1,000 examinees was generated. Three parameters for each item and two parameters for each examinee were used to generate these data. The values used for the item parameters are shown in Table 1. The examinee ability parameters were selected from a bivariate normal distribution with $\rho = 0$, $\mu = 0$, and Σ equal to the identity matrix.

The weight vectors used in this study were as follows. For an incorrect response, all of the weight vectors were set equal to zero. For a correct response, the following matrix and vectors were used:

$$\underset{\sim}{W} = (0,0) \qquad [45]$$

$$\underset{\sim}{U} = (1,0,0) \qquad [46]$$

$$\underset{\sim}{V} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \qquad [47]$$

As can be seen, only the first item parameter was selected by the weight vector, $\underset{\sim}{U}$, to act as the item difficulty parameter. The other two item parameters were selected by the weight matrix, $\underset{\sim}{V}$, to act as discrimination parameters. The resulting model is given by

Table 1
True and Estimated Parameters

Item	True Parameters 1	2	3	Estimates Unconditional 1	2	3	Conditional 1	2	3
1	-.65	1.50	.50	-.51	1.45	.44	-.48	1.27	.36
2	-1.40	.50	1.25	-1.34	.48	1.29	-1.56	.45	1.29
3	-.20	1.35	.15	-.02	1.49	.40	-.06	1.69	.25
4	.40	1.60	.55	.64	1.67	.84	.59	1.58	.63
5	.00	.50	1.15	.19	.48	1.02	.10	.51	1.04
6	-1.30	.35	1.05	-1.31	.50	1.17	-1.68	.42	1.25
7	.05	1.45	.35	.28	1.56	.43	.34	1.63	.35
8	.19	.25	1.40	.20	.27	1.18	.25	.20	1.23
9	-.17	.85	.85	-.02	1.07	.87	-.07	.89	.67
10	.14	1.75	.45	.40	1.60	.60	.34	1.45	.57
11	.37	.60	.80	.40	.66	1.02	.29	.61	.88
12	.87	1.65	.65	.82	1.49	.63	.70	1.38	.46
13	-.93	.35	1.35	-.98	.26	1.58	-1.00	.24	1.65
14	1.85	.65	1.65	1.80	.70	1.55	1.59	.54	1.29
15	.06	.65	.65	.16	.73	.66	.00	.80	.52
16	-.41	.45	1.45	-.31	.48	1.53	-.38	.23	1.60
17	-1.54	.75	1.25	-1.41	.72	1.32	-1.55	.71	1.14
18	.34	1.55	.25	.44	1.50	.32	.39	1.46	.29
19	-.15	.65	1.35	.14	.90	1.38	.04	.65	1.28
20	1.48	1.25	.45	1.54	1.41	.41	1.42	1.32	.22
21	-1.45	1.65	.45	-1.42	1.69	.46	-1.73	1.80	.48
22	.75	.45	1.35	.75	.45	1.42	.67	.40	1.35
23	-.75	.35	1.55	-.62	.33	1.75	-.63	.21	1.72
24	1.10	1.10	.30	.99	.92	.37	.93	1.11	.35
25	-.55	1.20	.15	-.42	1.35	.29	-.38	1.37	.24
26	.50	.50	1.00	.44	.66	1.03	.28	.49	1.12
27	-.15	1.45	.45	.10	1.36	.52	-.02	1.32	.42
28	.65	.70	.70	.78	.63	.66	.56	.68	.86
29	-1.00	1.00	.30	-1.02	1.16	.37	-.95	1.14	.38
30	1.00	.30	1.00	1.18	.27	1.27	.98	.27	1.14
31	-.25	.95	.25	-.16	.99	.20	-.04	1.11	.24
32	-.70	.15	1.50	-.65	.17	1.46	-.68	.03	1.51
33	.85	1.15	.45	1.10	1.31	.42	.98	1.29	.30
34	.05	.10	.95	-.02	.13	.92	.03	.20	1.09
35	-.95	1.35	.50	-.79	1.39	.64	-.80	1.28	.52
36	-1.50	.20	1.20	-1.45	.36	1.07	-1.38	.47	1.03
37	1.80	1.55	.55	2.16	1.84	.35	2.06	1.51	.44
38	-2.00	.15	1.15	-1.98	.32	1.05	-2.06	.30	1.09
39	-.90	1.40	.35	-.70	1.38	.22	-.70	1.18	.40
40	1.00	1.00	1.00	1.06	.92	.96	1.04	.87	.79
41	.15	1.25	.70	.26	1.33	.80	.28	1.17	.75
42	-1.50	.25	.95	-1.45	.40	.85	-1.60	.49	1.14
43	-1.25	.35	1.45	-1.02	.27	1.42	-1.13	.26	1.30
44	1.25	1.30	.25	1.37	1.25	.19	.95	1.26	.26
45	-2.00	1.15	.15	-1.89	.99	.09	-2.01	1.30	.13
46	1.75	.50	.50	1.80	.52	.59	1.65	.68	.64
47	.65	.65	1.30	.82	.48	1.39	.71	.49	1.27
48	-.25	1.00	.45	-.19	1.03	.59	-.17	1.03	.53
49	.35	.55	1.15	.40	.59	1.14	.24	.48	1.04
50	.00	.95	.15	-.05	.87	.25	-.07	1.11	.26

$$P(x|\underset{\sim}{\theta}_j,\underset{\sim}{\sigma}_i) = \frac{\exp[\sigma_{i1} + \sigma_{i2}\theta_{j1} + \sigma_{i3}\theta_{j2})x_{ij}]}{1 + \exp[\sigma_{i1} + \sigma_{i1}\theta_{j1} + \sigma_{i3}\theta_{j2}]} \ . \qquad [48]$$

Results

Table 1 shows the item parameter estimates obtained from both the conditional and unconditional item parameter estimation procedures. The estimates have be-n scaled to have the same means and standard deviations as the corresponding true item parameters. The correlations of the estimates with the true values are shown in Table 2. As can be seen, for these data there was very little difference in the quality of the estimates yielded by the two estimation procedures. Of course, this comparison is based on simulation data and on only one data set. Clearly, more research is needed before any definite conclusions about these procedures can be drawn.

Table 2
Intercorrelation Matrix for True and Estimated Item Parameters

	True			Estimation Procedure: Unconditional			Estimation Procedure: Conditional		
Parameter	σ_{T1}	σ_{T2}	σ_{T3}	$\hat{\sigma}_{U1}$	$\hat{\sigma}_{U2}$	$\hat{\sigma}_{U3}$	$\hat{\sigma}_{C1}$	$\hat{\sigma}_{C2}$	$\hat{\sigma}_{C3}$
σ_{T1}	1.00	.21	−.12	.99	.20	−.12	.99	.15	−.18
σ_{T2}	.21	1.00	−.75	.20	.97	−.72	.15	.96	−.79
σ_{T3}	−.12	−.75	1.00	−.12	−.72	.97	−.18	−.79	.96
$\hat{\sigma}_{U1}$	.99	.20	−.12	1.00	.24	−.13	.99	.19	−.20
$\hat{\sigma}_{U2}$	.20	.97	−.72	.24	1.00	−.73	.19	.96	−.80
$\hat{\sigma}_{U3}$	−.12	−.72	.97	−.13	−.73	1.00	−.20	−.80	.97
$\hat{\sigma}_{C1}$	.99	.15	−.18	.99	.19	−.20	1.00	.20	−.22
$\hat{\sigma}_{C2}$	.15	.96	−.79	.19	.96	−.80	.20	1.00	−.88
$\hat{\sigma}_{C3}$	−.18	−.79	−.96	−.20	−.80	.97	−.22	−.88	1.00

Discussion

The purposes of this paper were threefold. First, the fundamental concepts required when considering multidimensional models for the interaction of a person and a test item were defined. These concepts included the multidimensional latent space, the item difficulty function, and the item discrimination function. These definitions were conceived as multidimensional generalizations of similar concepts in unidimensional IRT models. Second, six existing multidimensional models were reviewed and, on the basis of their similarities, were classified into three general categories. The characteristics of these categories were described, and the general Rasch model was selected for further study on the basis of ease of parameter estimation. Third, estimation procedures for the parameters of the general Rasch model were described and applied to a set of simulation data that had been generated according to a two-dimensional special case of the model. The results indicated that a very close correspondence had been obtained between the estimated item parameters and those used to generate the simulation data.

On the basis of this information, two conclusions can be drawn. First, the concepts of difficulty and discrimination can be generalized to the multidimensional case, but the results are slightly different for compensatory and noncompensatory models. For the compensatory models, the item difficulty is defined by a linear function of the ability dimensions, while for the noncompensatory models, difficulty is defined by a vector of difficulty parameters. In both cases the slope of the item response surface at the difficulty function is a function of the discrimination parameters of the model.

A second conclusion that can be drawn is that the parameters of the general Rasch model can be estimated with acceptable accuracy for the simple two-dimensional case presented in the paper. This is, of course, very minimal evidence for the value of the estimation procedures; but combined with the work of Bock and Aitkin (1981), the results look fairly promising.

The results presented in this paper summarize only the initial steps in a thorough study of the applicability and usefulness of multidimensional latent trait models. Much further work needs to be done. The sample size requirements and the characteristics of the estimates obtained from the estimation procedures need to be determined. Procedures for determining the number of dimensions required for the multidimensional latent space must be developed as well as guidelines for interpreting the dimensions. The usefulness of the models for real data applications such as test construction and adaptive testing should be investigated. Also the advantages of these procedures over existing multidimensional procedures, such as factor analysis, should be studied.

REFERENCES

American College Testing Program. Content of the tests in the ACT Assessment. Iowa City IA: Author, 1980.

Andersen, E. B. A general latent structure model for contingency table data (Research Report No. 82). Copenhagen: Universitetets Statistiske Institut, 1982.

Andrich, D. A rating formulation for ordered response categories. Psychometrika, 1978, 43, 561-573.

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

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972, 37, 29-51.

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

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

Christoffersson, A. Factor analysis of dichotomized variables. Psychometrika, 1975, 40, 5-32.

Holzman, T. G., Glaser, R., & Pellegrino, J. W. Cognitive determinants of series completion: Individual and developmental differences. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Kim, J., & Mueller, C. W. Factor analysis: Statistical methods and practical issues. Beverly Hills CA: Sage Publications, 1978.

Kolakowski, D., & Bock, R. D. A Fortran IV program for maximum likelihood item analysis and test scoring: Normal ogive model (Research Memo No. 12). Chicago: The University of Chicago, Department of Education, Statistical Laboratory, 1970.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

Lord, F. M. Discussion: Session 2. In D. J. Weiss (Ed.), Proceedings of the 1977 Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

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

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

McKinley, R. L., & Reckase, M. D. Multidimensional latent trait models. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, March 1982.

Mulaik, S. A. A mathematical investigation of some multidimensional Rasch models for psychological tests. Paper presented at the annual meeting of the Psychometric Society, Princeton NJ, March 1972.

Muthén, B. Contributions to factor analysis of dichotomous variables. Psychometrika, 1978, 43, 551-560.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danske Paedagogike Institut, 1960.

Rasch, G. On general laws and the meaning of measurement in psychology. In J.

Neyman, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 4). Berkeley: University of California Press, 1961.

Reckase, M. D. Development and application of a multivariate logistic latent trait model (Doctoral dissertation, Syracuse University, 1972). Dissertation Abstracts International, 1973, 33. (University Microfilms No. 73-7762)

Reckase, M. D. Ability estimation and item calibration using the one and three parameter logistic models: A comparative study (Research Report 77-1). Columbia: University of Missouri, Department of Educational Psychology, November 1977.

Reckase, M. D. The formation of homogeneous item sets when guessing is a factor in item responses (Research Report 81-5). Columbia: University of Missouri, Department of Educational Psychology, August 1981.

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

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

Whitely, S. E. Multicomponent latent trait models for ability tests. Psychometrika, 1980, 45, 479-494.