DOCUMENT RESUME

ED 294 925 TM 011 689

AUTHOR	Batley, Rose-Marie; Boss, Marvin W.
TITLE	The Effects on Parameter Estimation of Correlated Abilities Using a Two-Dimensional, Two-Parameter Logistic Item Response Model.
PUB DATE	Apr 88
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Ability; *Computer Simulation; Equations (Mathematics); *Estimation (Mathematics); Foreign Countries; *Latent Trait Theory; *Multivariate Analysis; Vectors (Mathematics)
IDENTIFIERS	Item Parameters; *Parametric Analysis

ABSTRACT

The effects of correlated dimensions on parameter estimation were assessed, using a two-dimensional item response theory model. Past research has shown the inadequacies of the unidimensional analysis of multidimensional item response data. However, few studies have reported multidimensional analysis of multidimensional data, and, in those using simulated data, the results were usually based on replication. Multidimensional analysis of simulated two-dimensional item response data fitting the M2PL model of M. D. Reckase (1985) was done using the analysis program known as MIRTE. A Monte Carlo study was employed. Three data sets (2,000 ability vectors by 104 items) were generated to satisfy different degrees of correlation between the two abilities. The data sets and analyses were replicated 100 times each. Summary statistics on the 100 replications were used to examine the effects of the degree of correlation between ability dimensions. Results shed li on the degree of correlation between the two ability dimensions. Ability and item parameters were recovered well enough to encourage further investigation of and to justify limited use of multidimensional analysis. (Author/TJH)

ED294925

THE EFFECTS ON PARAMETER ESTIMATION OF CORRELATED ABILITIES USING A TWO-DIMENSIONAL, TWO-PARAMETER LOGISTIC ITEM RESPONSE MODEL.

Rose-Marie Batley

Marvin W. Boss

The University of Ottawa

Paper presented at the Annual Meeting of the

American Educational Research Association, New Orleans,

April 9, 1988

Abstract

The purpose of this study was to assess the effects of correlated dimensions on parameter estimation using a two-dimensional IRT model. Past research has shown the inadequacies of unidimensional analysis of multidimensional item response data. However, few studies have reported multidimensional analysis of multidimensional data and, in those which used simulated data, results were usually based on one replication.

Multidimensional analysis of simulated two-dimensional item response data fitting the M2PL model of Reckase (1985a) was done using the analysis program, MIRTE (Carlson, 1987).

Three data sets (2000 ability vectors by 104 items) were generated to satisfy different degrees of correlation between the two abilities. The three data sets and analyses were replicated 100 times each. Summary statistics on the 100 replications were used to assess the effects of degree of correlation between ability dimensions.

Results indicated that the degree of correlation between the two ability dimensions was less well recovered as $\rho(\theta_1,\theta_2)$ increased. Ability and item parameters were recovered well enough to encourage further investigation of and to justify limited use of multidimensional analysis.

3

Theoretical Framework

The original Item Response Theory (IRT) models were based on the assumption of unidimensionality (i.e., only one ability was required to correctly respond to all the items). When more than one ability accounts for test performance, the test is multidimensional and a Multidimensional Item Response Theory (MIRT) model is required to accurately fit the data.

Consider the situation in which items for a test are designed to measure one ability (e.g., mathematics) but require some amount of a second ability (e.g., verbal) in order to respond correctly. This second, required ability could be more crucial to success for some examinees than others. For example, students of English as a Second Language (ESL) may have sufficient mathematics ability but lack the required amount of verbal ability in order to make a correct response. It is reasonable to assume the two abilities are correlated to some extent. What happens to ability estimates if a MIRT model is used to fit the responses? How are the ability estimates affected by degree of correlation between the abilities?

Several authors (e.g., Ansley & Forsyth, 1985; Bogan & Yen; 1983; Dorans & Kingston, 1985; Drasgow & Parsons, 1983; McCauley & Mendoza, 1985; McKinley & Reckase, 1984; Reckase, 1979, 1985b; Reckase, Carlson, Ackerman & Spray, 1986) have considered the effects of analyzing known multidimensional data with a unidimensional item response model. The resulting estimates in most cases were not acceptable unless there was clearly one dominant dimension. Ansley and Forsyth (1985) reported that the unidimensional ability estimates were most highly related to the average of the multidimensional abilities. In the hypothetical educational situation described above, this would be unacceptable if students with high mathematics ability but low verbal ability were penalized in placement or selection procedures. Reckase et al (1986) found that the unidimensional ability estimates established from multidimensional data had different interpretations at different points on the unidimensional ability scale. By and large, the resulting unidimensional estimates from multidimensional data have been difficult to interpret and have not reflected well the original characteristics of the data.

In spite of findings that unidimensional models are not often robust to multidimensionality, few researchers have made use of multidimensional models to analyze multidimensional data. There are good reasons for this. Although MIRT models are being developed and tested, they are more complex than their unidimensional counterparts. Analysis of multidimensional data with multidimensional programs is expensive in terms of computer time. Few multidimensional analysis programs exist and none has undergone exhaustive testing. Only two programs have been readily available: (1) TESTFACT (Wilson, Wood & Gibbons, 1984); and (2) MAXLOG (McKinley & Reckase, 1983b). TESTFACT has been deemed inappropriate by some researchers because it uses a linear factor analytic procedure to describe the non-linear IRT relationship, a particularly contentious procedure with multidimensional data (Ansley, 1984; Lord, 1980; McDonald & Ahlawat, 1974; R. L. McKinley, personal communication, November 13, 1986). MAXLOG was written to provide parameter estimates for uncorrelated abilities. Results of pilot testing of a third multidimensional analysis program, MIRTE (Carlson, 1987), indicate that it estimates item parameters and abilities more efficiently and more accurately than MAXLOG and it can accommodate data from correlated dimensions. The program is designed to analyze data which fit the multidimensional two-parameter logistic ogive (M2PL) model (McKinley & Reckase, 1983a; Reckase, 1985b, 1986).

It is unreasonable to assume abilities are uncorrelated for most achievement tests. McKinley and Reckase (1984) considered the effects of analyzing data generated for correlated dimensions using MAXLOG. The ability and item estimates were confounded in the results of the data analysis. However, when the underlying abilities were correlated and a unidimensional analysis was used, again both unidimensional ability and item parameter estimates were affected (McKinley & Reckase, 1984).

Researchers who have used multidimensional analysis (e.g., McKinley, 1983; McKinley & Reckase, 1983a, 1983b, 1984; Muraki & Englehard, 1985) have indicated that a multidimensional model more adequately describes both real and simulated multidimensional data than does a unidimensional model. However, in most cases, the simulation studies have been based on no replications so that stability of estimates is difficult to determine. There is a need to know how consistently these estimates are recovered. The effects of both correlated abilities and differential secondary ability on parameter estimation need to be evaluated in a comprehensive, systematic manner.

Purpose of the Study

The purpose of this study was to determine the adequacy of multidimensional ability and item parameter estimates using a MIRT analysis. Specifically the question to be addressed is: what is the effect of correlated ability dimensions on parameter estimation for a two-parameter, two-dimensional IRT model?

Methodology

A Monte Carlo study was chosen to answer the research questions.

Model Description

The data for the study were generated to fit the multidimensional two-parameter logistic (M2PL) model (McKinley & Reckase, 1983a) which was updated by Reckase (1985b, 1986). A description of the updated version follows.

The mathematical formula is given by Equation (1).

$$P_{ij} = P(x_{ij} = 1 \mid \underline{a}_i, d_i, \underline{\theta}_j) = \frac{\exp(\underline{a}_i'\underline{\theta}_j + d_i)}{1 + \exp(\underline{a}_i'\underline{\theta}_j + d_i)}, \qquad (1)$$

$(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, N)$

6

where P_{ij} is the probability of a correct response to item i by examinee j; x_{ij} is the response (1 = correct; 0 = incorrect) of examinee j on item i; $\underline{a}_i$ is a vector of m discrimination parameters; d_i is a parameter representing the difficulty of item i; $\underline{\theta}_j$ is a vector of m ability parameters for individual j; N is the number of examinees; n is the number of items; and m is the number of dimensions.

This model is compensatory in that it allows high proficiency on one dimension to compensate for low proficiency on other dimensions in arriving at a correct response to a test item.

Reckase (1986) defined a multidimensional discrimination parameter for item i to be

$$MDISC_i = \left[\sum_{k=1}^{m} (a_{ik})^2 \right]^{0.5} , \qquad (2)$$

This parameter is related to the item characteristic curve on the multidimensional item response surface above the line through the origin of the ability space and to the point of maximum information and is therefore analogous to the unidimensional discrimination parameter (Carlson, 1987).

Reckase (1985b) also defined a multidimensional item difficulty parameter, $MDIF_i$, such that

$$MDIF_i = -d_i \Big/ \left[\sum_{k=1}^{m} (a_{ik})^2 \right]^{0.5} , \qquad (3)$$

$$= -d_i \Big/ MDISC_i$$

This parameter represents the distance between the origin of the m-dimensional ability space and the point in the space where the item information is a maximum. The line joining this point to the origin is at an angle of α_{ik} to the k^{th} ability dimension where

$$\cos \alpha_{ik} = a_{ik} \Big/ \left[\sum_{k=1}^{m} (a_{ik})^2 \right]^{0.5} , \qquad (4)$$

Program Description

The program used to analyze the two-dimensional data was MIRTE (Carlson, 1987). While written specifically to provide estimates of item and ability parameters for a M3PL model, the program is readily adapted to the M2PL model by letting the third item parameter equal zero. As well as estimation of abilities, item discriminations and difficulty, MIRTE provides estimates of standard errors for each of the parameter estimates. Estimates of the multidimensional item difficulty and discrimination are also provided. The method of estimation used is a variation of the joint maximum likelihood procedure using a modified Newton-Raphson iteration technique and the algorithm used is similar to that used in the unidimensional analysis program, LOGIST (Wingersky, Barton, & Lord, 1982). The MIRTE (version 2.00) used in this study was found to estimate parameters when dimensions were correlated better than MAXLOG (J. E. Carlson, personal communication, December, 1987). While MIRTE has been used in one recent study (Ackerman, 1987) to estimate item parameters, the author did not investigate questions considered in this study.

Data Description

Three different data sets were used (A1, A2, A3) representing cases in which both underlying abilities (θ_1 and θ_2) were normally distributed with mean 0, standard deviation 1. The difference among the thre sets was the degree of correlation between the abilities, namely 0.00, 0.25, and 0.50.

The simulated test consisted of 104 items, 26 items requiring only the first ability, 52 items requiring predominantly the first ability, 26 items requiring equal amounts of both abilities. A listing of the item parameters is provided in Table 1. Thirteen values of MDIF (ranging from -3 to +3 at intervals of 0.5) and two values of MDISC (2.00, 1.70) were chosen in order to cover the range of difficulties and to simulate realistic discrimination conditions in which the items were designed to discriminate well on the first ability. To meet the requirement that the items discriminate well on the first ability, four values of the angle, α_{i1}, (0°, 15°, 30°, 45°), were chosen. The discrimination indices, a_1 and a_2 (one for each dimension), were then generated to fit

Table 1. True Item Parameters for the 104 Items

α_{i1}	θ_i	d_i	Item	MDISC	a_{i1}	a_{i2}	Item	MDISC	a_{i1}	a_{i2}
0°	-3.0	-6	1	2.00	2.00	0.00	53	1.70	1.70	0.00
0°	-2.5	-5	2	2.00	2.00	0.00	54	1.70	1.70	0.00
0°	-2.0	-4	3	2.00	2.00	0.00	55	1.70	1.70	0.00
0°	-1.5	-3	4	2.00	2.00	0.00	56	1.70	1.70	0.00
0°	-1.0	-2	5	2.00	2.00	0.00	57	1.70	1.70	0.00
0°	-0.5	-1	6	2.00	2.00	0.00	58	1.70	1.70	0.00
0°	0.0	0	7	2.00	2.00	0.00	59	1.73	1.70	0.00
0°	0.5	1	8	2.00	2.00	0.00	60	1.70	1.70	0.00
0°	1.0	2	9	2.00	2.00	0.00	61	1.70	1.70	0.00
0°	1.5	3	10	2.00	2.00	0.00	62	1.70	1.70	0.00
0°	2.0	4	11	2.00	2.00	0.00	63	1.70	1.70	0.00
0°	2.5	5	12	2.00	2.00	0.00	64	1.70	1.70	0.00
0°	3.0	6	13	2.00	2.00	0.00	65	1.70	1.70	0.00
15°	-3.0	-6	14	2.00	1.932	0.518	66	1.70	1.642	0.44
15°	-2.5	-5	15	2.00	1.932	0.518	67	1.70	1.642	0.44
15°	-1.0	-4	16	2.00	1.932	0.518	68	1.70	1.642	0.44
15°	-1.5	-3	17	2.00	1.932	0.518	69	1.70	1.642	0.44
15°	-1.0	-2	18	2.00	1.932	0.518	70	1.70	1.642	0.44
15°	-0.5	-1	19	2.00	1.932	0.518	71	1.70	1.642	0.44
15°	0.0	0	20	2.00	1.932	0.518	72	1.70	1.642	0.44
15°	0.5	1	21	2.00	1.932	0.518	73	1.70	1.642	0.44
15°	1.0	2	22	2.00	1.932	0.518	74	1.70	1.642	0.44
15°	1.5	3	23	2.00	1.932	0.518	75	1.70	1.642	0.44
15°	2.0	4	24	2.00	1.932	0.518	76	1.70	1.642	0.44
15°	2.5	5	25	2.00	1.932	0.518	77	1.70	1.642	0.44
15°	3.0	6	26	2.00	1.932	0.518	78	1.70	1.642	0.44
30°	-3.0	-6	27	2.00	1.732	1.00	79	1.70	1.472	0.85
30°	-2.5	-5	28	2.00	1.732	1.00	80	1.70	1.472	0.85
30°	-2.0	-4	29	2.00	1.732	1.00	81	1.70	1.472	0.85
30°	-1.5	-3	30	2.00	1.732	1.00	82	1.70	1.472	0.85
30°	-1.0	-2	31	2.00	1.732	1.00	83	1.70	1.472	0.85
30°	-0.5	-1	32	2.00	1.732	1.00	84	1.70	1.472	0.85
30°	0.0	0	33	2.00	1.732	1.00	85	1.70	1.472	0.85
30°	0.5	1	34	2.00	1.732	1.00	86	1.70	1.472	0.85
30°	1.0	2	35	2.00	1.732	1.00	87	1.70	1.472	0.85
30°	1.5	3	36	2.00	1.732	1.00	88	1.70	1.472	0.85
30°	2.0	4	37	2.00	1.732	1.00	89	1.70	1.472	0.85
30°	2.5	5	38	2.00	1.732	1.00	90	1.70	1.472	0.85
30°	3.0	6	39	2.00	1.732	1.00	91	1.70	1.472	0.85
45°	-3.0	-6	40	2.00	1.414	1.414	92	1.70	1.202	1.202
45°	-2.5	-5	41	2.00	1.414	1.414	93	1.70	1.202	1.202
45°	-2.0	-4	42	2.00	1.474	1.414	94	1.70	1.202	1.202
45°	-1.5	-3	43	2.00	1.414	1.414	95	1.70	1.202	1.202
45°	-1.0	-2	44	2.00	1.414	1.414	96	1.70	1.202	1.202
45°	-0.5	-1	45	2.00	1.414	1.414	97	1.70	1.202	1.202
45°	0.0	0	46	2.00	1.414	1.414	98	1.70	1.202	1.202
45°	0.5	1	47	2.00	1.414	1.414	99	1.70	1.202	1.202
45°	1.0	2	48	2.00	1.414	1.414	100	1.70	1.202	1.202
45°	1.5	3	49	2.00	1.414	1.414	101	1.70	1.202	1.202
45°	2.0	4	50	2.00	1.414	1.414	102	1.70	1.202	1.202
45°	2.5	5	51	2.00	1.414	1.414	103	1.70	1.202	1.202
45°	3.0	6	52	2.00	1.414	1.414	104	1.70	1.202	1.202

the corresponding d and MDISC. The correlations between the original item parameters were: $\rho(d,a_1) = 0.004$; $\rho(d,a_2) = -0.004$; $\rho(a_1,a_2) = -0.738$; and $\rho(\text{MDIF},\text{MDISC}) = -0.002$.

9

Because of the dependency of a_1 and a_2, there is a larger correlation between these parameters. The same item parameters were used for each of the three data sets.

Procedure

The FORTRAN program M2PLGEN (Ackerman, 1985) was used to generate 2000 ability vectors (θ_1, θ_2) satisfying the distributions of θ_1 and θ_2 for Data Set A1. M2PLGEN uses a random seed and the IMSL (1979) subroutine GGNSM to generate random abilities. These ability vectors and the item parameters (a_1, a_2, d) were then used to generate response vectors (0s and 1s) for each of the 2000 simulees to each of the 104 items according to the M2PL model.

The 2000 x 104 matrix of response vectors was analyzed using MIRTE to provide estimates of θ_1, θ_2, a_1, a_2, d, MDIF, MDISC, α_1, and α_2. These results were filed. the random seed was incremented by two and the process was repeated. For Data Set A1 there were 100 replications. Summary statistics were calculated on the 100 replications.

This procedure was repeated for the other data set conditions. The same initial item parameter estimates for a_1 and a_2 were used for every replication in order to provide better control in the design. Finally, summary results from the three data sets were compared.

Results and Discussion

The purpose of this research was to determine the effects of correlated abilities on parameter estimation given a two-dimensional, two-parameter logistic item response model. First it should be determined if suitable ability data were generated to model the conditions specified. Then it needs to be determined whether MIRTE adequately estimated the parameters from the analysis of the response vectors generated.

Generation of (θ_1, θ_2): The ability data in all three data sets were generated to fit the specifications stated. The correlation between θ_1 and θ_2 for data generated over the 100 replications was recovered as -0.001 for Data Set A1, 0.251 for A2, and 0.500 for A3. The means for θ_1 and θ_2 were in the range 0 ± 0 004 and standard deviations were within 1 ± 0.003. There was very small variance (less than 0.0005) for these means and standard deviations in all data sets. There were no replications in which the ability data were not satisfactorily generated.

Recovery of Ability Parameters: In each of the three data sets over the 100 replications, $\hat{\theta}_1$ and $\hat{\theta}_2$ had means of 0.00 and standard deviations of 1.00. The standard deviation of the mean was less than 0.001 for all data sets. The recovery of these statistics is not particularly meaningful as a measure of accuracy in these cases because the MIRTE program rescales the theta estimates to mean 0, standard deviation 1 after each iteration in order to prevent drifting of the estimates.

In the data analysis, the program doesn't always identify dimensions one and two correctly. In order to avoid confusing the dimensions during the 100 replications, a check was made during each data analysis on the first thirteen item discrimination parameter estimates. (These items were pure on θ_1.) If the sum of the first thirteen a_1 estimates was less than the sum of the first thirteen a_2 estimates, the estimations for the dimensions were flipped.

The mean average absolute deviation of $\hat{\theta}_1$ from the true θ_1 (AAD($\hat{\theta}_1$)) ranged from 0.44 to 0.46 (see Table 2). Increasing $\rho(\theta_1, \theta_2)$ did not appear to affect this. The mean average absolute deviation of $\hat{\theta}_2$ (AAD($\hat{\theta}_2$)) ranged from 0.54 to 0.41 and seemed to be more affected by the correlation between the abilities. As $\rho(\theta_1, \theta_2)$ increased, the AAD($\hat{\theta}_2$) decreased. This could be a result of the compensatory nature of the M2PL model. There was very little variance over replications in these AADs (0.001 for $\hat{\theta}_1$; 0.002 for $\hat{\theta}_2$) so that the thetas appear to have been recovered consistently across the three data sets.

As θ_1 and θ_2 became more highly correlated, θ_2 appeared to be better estimated (AAD($\hat{\theta}_2$) decreased). This was supported by the mean correlation between θ_2 and $\hat{\theta}_2$.

As $\rho(\theta_1, \theta_2)$ increased, θ_2 became more highly correlated with θ_2 (Table 2). In all three data sets, θ_2 was recovered fairly well according to $r(\theta_2, \hat{\theta}_2)$. The mean standard error of the thetas (as calculated by MIRTE) was approximately 0.259, almost half the size of the AADs. The variance in these mean standard errors was very small although the standard errors were more spread out as the correlation between the dimensions increased.

Table 2
Mean Values of Statistics for Estimated Thetas (over 100 replications)

Data Set	$\rho(\theta_1,\theta_2)$	$AAD(\hat{\theta}_1)$	$AAD(\hat{\theta}_2)$	$r(\hat{\theta}_1,\hat{\theta}_2)$	$r(\theta_1,\hat{\theta}_1)$	$r(\theta_2,\hat{\theta}_2)$	$r(\theta_1,\hat{\theta}_2)$	$r(\theta_2,\hat{\theta}_1)$
A1	0.00	0.441	0.544	0.062	0.842	0.764	0.505	-0.295
A2	0.25	0.446	0.470	0.179	0.842	0.824	0.603	-0.050
A3	0.50	0.459	0.412	0.282	0.831	0.865	0.699	-0.209

The ability θ_1 was also well recovered as $r(\theta_1, \hat{\theta}_1)$ was greater than 0.83 for all three data sets. In Data Set A3, θ_2 appeared to be recovered better than θ_1 in spite of the fact that few items were measuring the θ_2-space. This was also supported by the decreasing $AAD(\hat{\theta}_2)$ as the correlation between the ability dimensions increased. As $\rho(\theta_1, \theta_2)$ increased, θ_1 was less well recovered but θ_2 was better recovered.

The correlation between the ability vectors was not well recovered. As $\rho(\theta_1, \theta_2)$ increased, MIRTE tended to produce ability estimates which were less correlated than the generated abilities. The difference between $\rho(\theta_1, \theta_2)$ and $r(\hat{\theta}_1, \hat{\theta}_2)$ increased as $\rho(\theta_1, \theta_2)$ increased. This result agrees with that reported by Carlson (1987)

Recovery of Item Parameters: In the maximum likelihood estimation procedures used in MIRTE, ability estimates are used to improve item parameter estimates and vice versa. Hence, the final estimates are affected by each other. As $\rho(\theta_1, \theta_2)$ increased, what happened to the item parameter estimates?

Statistics on the item difficulty parameters are summarized in Table 3. In all three data sets, $r(d, \hat{d}) = 0.997$ indicating good estimation of the item difficulty. As $\rho(\theta_1, \theta_2)$ increased, the mean and standard deviation of $\hat{d}$ increased slightly but remained close to the original parameter statistics. The AAD($\hat{d}$) increased slightly as the correlation between the ability dimensions increased indicating that $\underline{d}$ was being less well recovered. The mean and standard deviation of the multidimensional difficulty parameter, MDIF, were recovered well although here again MDIF was less well recovered as $\rho(\theta_1, \theta_2)$ increased.

Table 3
Summary of Mean Statistics for Item Difficulty (over 100 replications)

Data Set	d	s(d)	se($\hat{d}$)	AAD($\hat{d}$)	MDIF	s(MDIF)	r(d,$\hat{d}$)	r(MDIF,$\widehat{\text{MDIF}}$)
True	0.009	3.771	-----	------	-0.005	2.058	-----	-----
A1	0.009	3.929	0.112	0.224	0.006	2.079	0.997	0.995
A2	0.010	3.936	0.109	0.228	0.005	2.028	0.997	0.994
A3	0.030	3.936	0.106	0.232	0.012	1.999	0.997	0.991

se - standard error from MIRTE program

Discrimination parameter estimates have been reported to be affected more by multidimensional data and analysis. This result was also evident in this study. The mean of $\hat{a}_1$ was lower than the true mean and the standard deviation was higher than the true standard deviation for all three data sets (see Table 4). The mean of $\hat{a}_2$ was much higher than the true mean of 0.678. In fact the mean of $\hat{a}_2$ was higher than the mean estimates of a_1 and approached the true mean of a_1 as $\rho(\theta_1, \theta_2)$ increased. Both means increased slightly as $\rho(\theta_1, \theta_2)$ increased. The standard deviation of $\hat{a}_2$ was higher than the true standard deviation but there was not as large a difference here as with $\hat{a}_1$. Standard errors of estimation of $\hat{a}_1$ and $\hat{a}_2$ were approximately 0.09 but the AADs were much larger, particularly for $\hat{a}_2$. As the correlation between the two ability dimensions increased, the AAD($\hat{a}_2$) increased slightly indicating a_2 was being

13

less well recovered. The AAD($\hat{a}_1$) was approximately 0.5 for all three data sets. The estimates of a_1 and a_2 ranged over a much wider interval than the originals. Although in all three data sets the lower bound (0.010) was hit for some of both a_1 and a_2 estimates, the upper bound of 4.500 was not hit until Data Set A3.

Table 4
Summary of Mean Statistics for Item Discrimination (over 100 replications)

Data Set	a_1	$s(a_1)$	$se(\hat{a}_1)$	AAD($\hat{a}_1$)	a_2	$s(a_2)$	$se(\hat{a}_2)$	AAD($\hat{a}_2$)	MDISC	s(MDISC)
True	1.637	0.251	-----	-----	0.678	0.496	-----	-----	1.850	0.151
A1	1.195	0.569	0.099	0.500	1.379	0.512	0.096	0.707	1.957	0.288
A2	1.201	0.528	0.095	0.486	1.448	0.582	0.094	0.775	2.013	0.319
A3	1.202	0.502	0.093	0.490	1.510	0.628	0.093	0.836	2.057	0.381

The multidimensional discrimination parameter, MDISC, was recovered with a higher mean and higher standard deviation in all three data sets. There appears to be a rotational indeterminacy in the recovery of the discrimination parameters and a tendency to spread the discrimination parameter estimates over the entire space even though they originally did not cover the entire space.

This was supported by the statistics on the angle estimates, $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Originally α_1 had a mean of 22.50°. This was recovered in all data sets at over 49°. Similarly, α_2, whose original mean was 67.50°, was recovered in all data sets at just over 40°. The original standard deviation of 16.85° increased for the estimates to approximately 20°. There seemed to be an attempt to cover the entire $\theta_1\theta_2$-space in estimation of parameters related to discrimination. Estimates of α_1 and α_2 ranged from very close to 0° to almost 90°.

Correlation coefficients again were used to determine adequacy of parameter recovery (Table 5). In all cases, a_1 correlated more highly with $\hat{a}_1$ than with $\hat{a}_2$. Similarly, a_2 correlated more highly with $\hat{a}_2$ than it did with $\hat{a}_1$. As well, a_2

correlated higher with $\hat{a}_2$ than a_1 did with $\hat{a}_2$. The anomaly in the correlations was that a_1 correlated less highly with $\hat{a}_1$ than a_2 did with $\hat{a}_1$. As the discrimination paran 'ers appear to be dispersed across the $\theta_1\theta_2$-space, this may account for the apparent better recovery of a_2 than of a_1. That the standard deviation of a_2 was twice as large as that of a_1 may also account for the higher correlations of both a_1 and a_2 with $\hat{a}_2$. The greater variability in a_2 would allow for higher correlations.

Table 5
Mean Correlations for Item Discrimination Values (over 100 replications)

Data Set	$r(a_1,\hat{a}_1)$	$r(a_2,\hat{a}_2)$	$r(\hat{a}_1,\hat{a}_2)$	$r(a_1,\hat{a}_2)$	$r(a_2,\hat{a}_1)$
A1	0.834	0.893	-0.765	-0.572	-0.865
A2	0.818	0.899	-0.769	-0.587	-0.830
A3	0.760	0.895	-0.735	-0.587	-0.747

The correlation between $\hat{a}_1$ and $\hat{a}_2$ was slightly stronger than the true parameter correlation of -0.738 except in the Data Set A3 where it was slightly smaller. The multidimensional discrimination parameter, MDISC, did not correlate as highly with its estimate. This correlation was highest (0.600) when the ability dimensions were uncorrelated and decreased as the correlation between the abilities increased.

The correlation between d and a_1 was 0.004. This was recovered as 0.019, 0.030, and 0.056 for Data Sets A1, A2, and A3 respectively. The correlation between d and a_2 was -0.004 which was recovered as 0.004, -0.011 and 0.004 for the three data sets respectively. This would suggest again that the second dimension is being better estimated. This could be partially a result of the items of the simulated test not covering the entire latent space and the variability in a_2 being so much greater than in a_1 However, the AAD($\hat{a}_2$) did not support the conclusion that a_2 is better recovered than a_1. The correlation between MDISC and MDIF was -0.002 and was recovered as -0.031, -0.027, and -0.073 for the three data sets respectively. There was no evident trend here.

15

Conclusions

The purpose of the study was to determine the effects of correlated dimensions in multidimensional estimation of data to fit the M2PL model.

Results of this research support the use of MIRTE to analyze two-dimensional data fitting the M2PL model where the ability dimensions are normally distributed over a full range. Even as the correlation between the ability dimensions increased from 0.00 to 0.50, both item and ability parameters seemed to be recovered well enough to justify the use of a multidimensional analysis program rather than forcing unidimensionality by using one of the unidimensional analysis programs available. Correlation coefficients between parameters and corresponding estimates were high.

Average absolute deviations for $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{a}_1$ were of a similar size. The AAD for $\hat{a}_2$ was much larger indicating the second dimension was not being as accurately recovered. As there were fewer items measuring the second dimension, recovery would be more difficult. The a_1 discrimination parameters were slightly underestimated while the a_2 discrimination parameters were overestimated.

Item difficulty parameters have traditionally been recovered better than other parameters and results of this research continued to support this phenomenon. The AAD($\hat{d}$) was much smaller than for the other variables and the correlations between the difficulty parameters and estimates all had $|r| > 0.98$. It should be remembered that that during the estimation procedures, to estimate item parameters there were 2000 ability vectors, while the estimation of the ability vectors was done from estimates for only 104 items. This would help account for better estimation of the single difficulty parameter.

Both theta estimates were accurately estimated in terms of correlation. However, average absolute deviations of 0.44 to 0.46 for $\hat{\theta}_1$ and 0.41 to 0.54 for $\hat{\theta}_2$ depending on the correlation between the ability dimensions were found. The correlation between the ability estimates for the two dimensions was less well recovered as the correlation between the ability dimensions increased. The correlation between the theta estimates was lower than the correlation between the true thetas. This was in agreement with results reported by Carlson (1987).

It would be useful to know how great this reduction of the correlation coefficient becomes as $\rho(\theta_1, \theta_2)$ increased beyond 0.50. As well, further research should consider the effects on parameter estimation when α_1, α_2, a_1, and a_2 are generated to cover the entire space. It is possible that the anomalies associated with the recovery of a_2 would be less severe given a different set of item characteristics. Conclusions based on the estimates of parameters related to the second dimension in particular are tentative as it appears there was an attempt by MIRTE to estimate over the entire latent space. This seemed to have affected estimates on the second dimension moreso than those for the first ability dimension.

Increasing the correlation between the true thetas did affect the estimation of parameters. The important parameter least well recovered in this study seemed to be a_2. This could be explained by the fact that there were fewer items measuring the second dimension. There is cause for concern over the size of the AADs for ability and discrimination parameter estimates in all three data sets.

The results of this research indicate a good future for the application of multidimensional models using the estimation procedures of MIRTE but additional studies are needed.

References

Ackerman, T. A. (1985). *M2PLGEN: A computer program for generating thetas and response strings corresponding to the M2PL model.* Iowa City, Iowa: American College Testing.

Ackerman, T. A. (1987, April). *The use of unidimensional item parameter estimates of multidimensional items in adaptive testing.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Ansley, T. N. (1984). An empirical investigation of the effects of applying a unidimensional atent trait model to two-dimensional data. (Doctoral dissertation, University of Iowa, 1984). *Dissertation Abstracts International, 45/07, 2074A.*

Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9(1),* 37-48.

Bogan, E. D., & Yen, W. M. (1983, April). *Detecting multidimensionality and examining its effects on vertical equating with the three-parameter logistic model.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec.

Carlson, J. E. (1987). *Multidimensional Item Response Theory Estimation: A Computer Program* (ACT Research Report 87-19). Iowa City, IO: American College Testing Program.

Dorans, N. J., & Kingston, N. M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE verbal scale. *Journal of Educational Measurement, 22(4),* 249-262.

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

International Mathematical and Statistical Libraries (1979). IMSL Library (7th ed.) Houston, TX: Author.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

McCauley, C. D., & Mendoza, J. (1985). A simulation study of item bias using a two-parameter item response model. *Applied Psychological Measurement, 9(4),* 389-400.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82-99.

McKinley, R. L. (1983, April). *A multidimensional extension of the two-parameter logistic latent trait model.* Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec. (ERIC Document Reproduction Service No. ED 228 326)

McKinley, R. L., & Reckase, M. D. (1983a). *An application of a multidimensional extension of the two-parameter logistic latent trait model* (ONR-83-3). (ERIC Document Reproduction Service No. ED 240 168)

McKinley, R. L., & Reckase, M. D. (1983b). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. *Behavior Research Methods and Instrumentation, 15(3),* 389-390.

McKinley, R. L., & Reckase, M. D. (1984). *An investigation of the effect of correlated abilities on observed test characteristics* (Research Report). Iowa City, Iowa:

American College testing Program, Test Development Division: . (ERIC Document Reproduction Service No. ED 249 249)

Muraki, E., & Englehard, G. (1985). Full-information item factor analysis: applications of EAP scores. *Applied Psychological Measurement, 9(4)*, 417-430.

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

Reckase, M. D. (1985a). *Models for multidimensional tests and hierarchical structured training materials* (Research Report ONR-85-1). Iowa City, Iowa: American College Testing Program.

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

Reckase, M. D. (1986, April). *The discriminating power of items that measure more than one dimension.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Reckase, M. D., Carlson, J. E., Ackerman, T. A., & Spray, J. A. (1986, June). *The interpretation of unidimensional IRT parameters when estimated from multidimensional data.* Paper presented at the annual meeting of the Psychometric Society, Toronto, Ontario.

Wilson, D., Wood, R. L., & Gibbons, R. (1984). *TESTFACT: Test Scoring and item factor analysis.* [Computer program]. Moorsville, IN: Scientific Software, Inc.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide.* Princeton, NJ: Educational Testing Service.