DOCUMENT RESUME

ED 258 992 TM 850 360

AUTHOR Doody, Evelyn N.
TITLE Examining the Effects of Multidimensional Data on Ability and Item Parameter Estimation Using the Three-Parameter Logistic Model.
PUB DATE Mar 85
NOTE 70p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Computer Simulation; Estimation (Mathematics); Factor Structure; Item Analysis; *Latent Trait Theory; Mathematical Models; Multidimensional Scaling; Research Methodology; Statistical Analysis; Test Items
IDENTIFIERS Item Parameters; Robustness; *Three Parameter Model; *Unidimensional Scaling

ABSTRACT

The effects of varying degrees of correlation between abilities and of various correlation configurations between item parameters on ability and item parameter estimation using the three parameter logistic model were examined. Ten two-trait configurations and one unidimensional test configuration for 30 item tests were simulated. Each configuration consisted of a specific item parameter configuration and a specific correlation between traits on two dimension. Six conditions were simulated for each configuration, ranging from an easy to a hard test. The accuracy of item and ability parameter estimation was examined using correlations; KR-20 coefficients and factor analyses were also performed. The factor analyses supported a division of the simulated multidimensional data sets into groups according to the discrimination parameter "loads" on the two dimensions. The tests either both load heavily on both dimensions (both tests are multidimensional), one test loads heavily on one dimension and the other loads heavily on the same dimension (both tests are unidimensional), one test is unidimensional and one is multidimensional, or one test loads heavily on one dimension and the other test loads heavily on the other dimension. Results indicated the poorest item parameter estimations occur for the situation in which one test is unidimensional and one is multidimensional. (Author/DWH)

EXAMINING THE EFFECTS OF MULTIDIMENSIONAL DATA ON ABILITY AND ITEM PARAMETER ESTIMATION USING THE THREE-PARAMETER LOGISTIC MODEL

Evelyn N. Doody
CTB/McGraw-Hill

Presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, April, 1985.

ABSTRACT

The effects of varying degrees of correlation between abilities and of various correlation configurations between item parameters on ability and item parameter estimation using the three-parameter logistic model was examined. Ten two-trait and one unidimensional test configurations for thirty item tests were simulated for 6000 simulees. Each configuration consists of a specific item parameter configuration and a specific correlation between traits on two dimensions. Six conditions were simulated for each configuration, ranging from a very easy to a very hard test. The accuracy of item and ability parameter estimation was examined using correlations; KR-20 coefficients and factor analyses were also performed. The factor analyses supported a division of the simulated multidimensional data sets into groups according to how the discrimination parameter "loads" on the two dimensions. The tests either both load heavily on both dimensions, (both tests are multidimensional), one tests loads heavily on one dimension and the other loads heavily on the same dimension (both tests are unidimensional), one test is unidimensional and one is multidimensional, or one test loads heavily on one dimension and the other test loads heavily on the other dimension. The results indicate that the poorest item parameter estimations occur for the situation in which one test is unidimensional and one is multidimensional.

EXAMINING THE EFFECTS OF MULTIDIMENSIONAL DATA ON ABILITY AND ITEM PARAMETER ESTIMATION USING THE THREE-PARAMETER LOGISTIC MODEL

INTRODUCTION

Several multidimensional models have been proposed and some research has been conducted using these models (Doody-Bogan & Yen, 1983; Hattie, 1982; McKinley, 1983; McKinley & Reckase, 1982, 1983a, 1983b, 1984; Reckase, 1979; Reckase & McKinley, 1982). However, use of these models has not yet proven feasible.

Most of the item response theory (IRT) methodology that has been developed is applicable only to the limited case of one-dimensional data, in which case the assumption of unidimensionality is required in order to estimate the item and ability parameters. Unfortunately, since in most practical applications that assumption is not realistic, and useful multidimensional estimation procedures are not yet available, practitioners must either fall back on traditional methodology or inappropriately apply IRT methodology while hoping for robustness to violation of the unidimensionality assumption. Such robustness remains undemonstrated.

Violation of the unidimensionality assumption has been suggested as a problem in estimation of item parameters (Loyd & Hoover, 1980; Cook & Eignor, 1981). It is informative to examine the effect of violation of the unidimensionality assumption on the estimation of the item parameters, a, b, and c, and on the estimation of ability,

θ. This effect can then be considered when parameters are estimated in situations in which multidimensionality is known or suspected and no usable estimation procedure for multidimensional data can be found.

Test Analysis Model

The statistical model to be used in this study for analyzing item responses is the three-parameter logistic model. This model assumes that an individual's performance on a test is influenced by only one important unobservable characteristic, θ, which is called a (latent) trait or ability. The three-parameter logistic model assumes that the probability of a correct response to item i by person j with ability level, θ, is:

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp(-1.7a_i(\theta_j - b_i))} \qquad (1)$$

where a_i, b_i, and c_i are the discriminating power, difficulty, and lower asymptote or guessing parameter of item i, respectively.

The accuracy of item parameter estimation is affected by several things, including the accuracy of the estimation program (McKinley & Reckase, 1980), the size of the calibration sample (Hambleton, Swaminathan, Cook, Eignor, & Gifford, 1978; Reckase, 1977), and the percent of test variance accounted for by the first factor found when the data are factor analysed (Reckase, 1979). Violation of the unidimensionality assumption has been suggested as a problem in estimation of item parameters (Loyd & Hoover, 1980; Cook & Eignor, 1981).

Objectives

The purpose of this research is to investigate the robustness of item and ability parameter estimation using the three-parameter logistic model to violation of the unidimensionality assumption, and to examine the effects of specific multidimensional data configurations on parameter estimation using the three-parameter logistic model.

Educational or Scientific Importance of the Study

Most commonly used IRT models assume unidimensionality. However, this assumption is not strictly satisfied by item pools in most practical situations (Lord, 1968). While the assumption of unidimensionality is acceptable in the case of aptitude tests, that assumption is unrealistic for many tests, including most achievement tests (McKinley & Reckase, 1982; Reckase, 1979, 1981; Sympson, 1978). Any factor that influences an examinee's score on a test, other than the one latent trait (ability) assumed for the one-dimensional model, will violate the assumption of unidimensionality. Guessing, speededness, fatigue, cheating, random answering, or accidently overlooking or skipping an item are possible factors. The existence of two or more cognitive traits is one such possible factor. An achievement test in mathematics might require both reading skill and mathematical reasoning. An achievement test in science might require both reading and knowledge of science facts. If so, the assumption of unidimensioality does not appear to be met. Nevertheless, IRT methodology has well-known advantages over traditional methodology and is applied in situations where it may not be appropriate. Hence it is informative to determine the effects of multidimensionality on parameter estimation. It is equally important to develop guidelines for educators and researchers concerned with achievement testing, who wish to benefit from the advantages of IRT methodology.

METHOD

The simulations begin with the generation of two-dimensional data sets from a multidimensional model using prespecified parameters through an investigation into the effects on parameter estimation of various multidimensional conditions, and end with a re-examination of the accuracy of the parameter estimation through cross-validation.

Data Generation

The main question to be examined in this research is how robust parameter estimation based on the three-parameter logistic model is to violation of the unidimensionality assumption underlying the estimation. The unidimensionality assumption is violated whenever the scores that are being equated are multidimensional in the sense that an examinee's score on a test is the result of more than one latent trait. The data can be the result of more than one latent trait and can also vary in the degree of correlation that exists between these traits. Since infinitely many multidimensional data sets fulfilling these requirements are possible, this research project is necessarily limited to a few of the possibilities.

Number of dimensions and degrees of correlation. The two-dimensional case was chosen for this research as typical of published tests and as a starting point in examining the robustness of parameter estimation to violation of the unidimensionality assumption. Examining all possibilities is beyond the scope of this research.

The choice of correlations was limited to that which seemed possible for a published test, the Comprehensive Tests of Basic Skills, Forms U and V (CTBS/U, CTBS/V; CTB/McGraw-Hill, 1981). A correlation of zero was chosen to simulate data sets on which Trait 1 and Tr[illegible] 2 have no correlation. Correlations of .3 and .[illegible] [illegible] chosen to simulate data sets on which Trait 1 and Trait 2 have low correlation, correlations of .5, .6, and .75 were used to simulate two traits that are more highly correlated, and .9 was used to simulate two traits that are highly correlated. One unidimensional data set, which is representative of a situation in which the correlation between traits is 1.0 was also generated to be used as a criterion against which the analyses of the multidimensional data sets could be compared.

Multivariate model. Two-dimensional data sets were generated using the multidimensional model described by Doody-Bogan and Yen (1983). This model is an extension of the three-parameter logistic latent-trait model. The multivariate logistic model is:

$$P_i(\theta_{jt}) = c_i + \frac{(1 - c_i)}{1 + \exp(-1.7 \sum_{t=1}^{m} a_{it}(\theta_{jt} - b_{it}))} \quad (2)$$

where $P_i(\theta_{jt}) = P_i(\theta_{j1},\theta_{j2},\ldots,\theta_{jm})$, the probability of a correct response to item i by a person j whose location in an m-dimensional latent space is described by abilities $\theta_{j1}, \theta_{j2},\ldots,\theta_{jm}$; θ_{jt} represents the ability of person j on trait t, a_{it} is the discrimination of item i with respect to latent trait t, b_{it} is the difficulty of item i with respect to latent trait t, and c_i is the guessing parameter for item i. Note that when m = 1, this model reduces to the univariate logistic three-parameter model of Birnbaum (1968). The model was used with m = 1 to simulate the unidimensional data. Thirty item tests were used.

Item parameter values and item pool. Discrimination, difficulty, and guessing values were chosen as in Doody-Bogan and Yen (1983). The base pool consists of 30 items. Since the existence of two traits is assumed, two discrimination parameters, two difficulty parameters, and one guessing parameter per item are required. Two test levels were simulated, a harder test (Test 2) and an easier test (Test 1). Item parameters for the harder test were estimated using simulees with higher ability levels than those used to estimate item parameters for the easier test.

Six data sets were simulated per data configuration, ranging from a very easy test to a very hard test. For different levels of difficulty, 1.0 was added to the base b_i values to simulate the hardest test and 1.0 was subtracted from the base b_i values to simulate the easiest test. Similarly, .5 was subtracted from and added to the base values to simulate slightly different levels of difficulty. These differences in difficulty are represented as $\bar{b}_2 - \bar{b}_1 = 0.0$, 1.0, and 2.0, where b_2 is the mean difficulty of the harder test (Test 2) and b_1 is the mean difficulty of the easier test (Test 1).

For each test configuration, item parameters a_{i1}, b_{i1}, a_{i2}, b_{i2}, and c_i were randomly assigned to Traits 1 and 2 for both Tests 1 and 2, with the restriction that the desired correlations between parameters were approximated as closely as possible. Randomizations were tried until the desired correlations were obtained.

Test configurations. Ten two-trait data test configurations were chosen to be simulated. Each was chosen as being typical of a possible achievement test. Table 1 shows the desired trait and item parameter correlations for the simulated data sets. Table 2 shows possible tests where such correlations might exist. Table 3 shows the item parameters used for each data configuration.

Table 1

Correlations for Simulated Data Sets

Configuration	Lower Level: Trait 1	Lower Level: Trait 2	Upper Level: Trait 1	Upper Level: Trait 2
1	$r(\theta_1,\theta_2) = 0$		same as	
	$r(a_1,b_1) = 0$	$r(a_2,b_2) = 0$	lower	
	$r(a_1,a_2) = 0$	$r(b_1,b_2) = 0$	level	
2	$r(\theta_1,\theta_2) = .3$		$r(\theta_1,\theta_2) = .3$	
	$r(a_1,b_1) = -.3$	$r(a_2,b_2) = 0$	$r(a_1,b_1) = -.3$	$r(a_2,b_2) = 0$
	$r(a_1,a_2) = 0$	$r(b_1,b_2) = 0$	$r(a_1,a_2) = .2$	$r(b_1,b_2) = 0$
3	$r(\theta_1,\theta_2) = .4$		same as	
	$r(a_1,b_1) = .2$	$r(a_2,b_2) = .5$	lower	
	$r(a_1,a_2) = 0$	$r(b_1,b_2) = 0$	level	
4	$r(\theta_1,\theta_2) = .5$		$r(\theta_1,\theta_2) = .5$	
	$r(a_1,b_1) = .4$	$r(a_2,b_2) = 0$	$r(a_1,b_1) = .4$	$r(a_2,b_2) = .7$
	$r(a_1,a_2) = 0$	$r(b_1,b_2) = 0$	$r(a_1,a_2) = 0$	$r(b_1,b_2) = .5$
5	$r(\theta_1,\theta_2) = .5$		same as	
	$r(a_1,b_1) = 0$	$r(a_2,b_2) = 0$	lower	
	$r(a_1,a_2) = -.8$	$r(b_1,b_2) = 0$	level	
6	$r(\theta_1,\theta_2) = .5$		same as	
	$r(a_1,b_1) = -.3$	$r(a_2,b_2) = 0$	lower	
	$r(a_1,a_2) = -.5$	$r(b_1,b_2) = 0$	level	
7	$r(\theta_1,\theta_2) = .6$		same as	
	$r(a_1,b_1) = 0$	$r(a_2,b_2) = 0$	lower	
	$r(a_1,a_2) = 0$	$r(b_1,b_2) = 0$	level	
8	$r(\theta_1,\theta_2) = .75$		same as	
	$r(a_1,b_1) = .4$	$r(a_2,b_2) = 0$	lower	
	$r(a_1,a_2) = 0$	$r(b_1,b_2) = 0$	level	

(table continues)

Table 1 (continued)

Configuration	Lower Level Trait 1	Lower Level Trait 2	Upper Level Trait 1	Upper Level Trait 2
9	$r(\theta_1,\theta_2)$ = .9		same as lower level	
	$r(a_1,b_1)$ = 0	$r(a_2,b_2)$ = 0		
	$r(a_1,a_2)$ = 0	$r(b_1,b_2)$ = 0		
10	$r(\theta_1,\theta_2)$ = .9		same as lower level	
	$r(a_1,b_1)$ = 0	$r(a_2,b_2)$ = 0		
	$r(a_1,a_2)$ = .8	$r(b_1,b_2)$ = .8		
U	unidimensional		unidimensional	

Note: All item parameters are written without the item subscript, i.
All ability parameters are written without the person subscript, j.

Table 2

Applications to Real Data

Test Name	Test 1 Trait 1	Test 1 Trait 2	Test 2 Trait 1	Test 2 Trait 2
1. Language Mechanics	end punctuation	middle punctuation	end punctuation	middle punctuation
2. Mathematics Computation	other items	decimals*	other items	decimals
3. Mathematics Concepts and Applications	other items	fractions & conversions	other items	fractions & conversions
4. Mathematics Computation	other items	fractions*	other items	fractions
5. Social Studies	reading	graphs	reading	graphs
6. Mathematics Computation	other items	decimals	other items	decimals
7. Science	reading	science facts	reading	science facts
8. Mathematics Concepts and Applications	mathematics	reading	mathematics	reading
9. Language Mechanics	reading	punctuation	reading	punctuation
10. Reading Comprehension	reading	vocabulary	reading	vocabulary

* not measured at this level.

Table 3

Base Item Parameters

Configuration	Easy Test: AL1	AL2	BL1	BL2	CL	Hard Test: AH1	AH2	BH1	BH2	CH
1	0.80	0.80	0.68	-1.25	0.18	0.90	1.00	0.62	-2.00	0.16
	2.00	0.80	0.18	1.52	0.20	0.90	1.00	-0.03	0.04	0.18
	1.00	2.00	0.62	-0.38	0.18	2.00	1.40	-0.31	-0.52	0.17
	1.10	1.00	-0.17	-2.00	0.20	1.00	1.10	1.35	0.49	0.20
	0.50	0.90	0.32	-0.03	0.16	0.60	1.80	1.02	0.18	0.19
	1.20	1.80	0.40	0.74	0.20	1.00	2.00	0.95	-1.25	0.20
	1.00	1.00	0.11	0.18	0.16	0.80	1.10	-0.84	0.53	0.16
	1.20	0.50	-0.66	0.62	0.16	1.00	0.50	-2.00	-0.10	0.18
	1.20	1.20	-0.31	0.11	0.18	1.10	0.90	0.26	-0.31	0.20
	0.90	0.80	-0.84	0.32	0.20	0.60	1.10	-0.10	0.32	0.22
	1.10	1.20	-2.00	-0.84	0.19	1.10	0.60	0.32	0.11	0.22
	0.60	0.70	-0.45	-1.46	0.19	1.00	1.00	-0.45	-0.84	0.20
	0.80	0.60	0.49	-0.17	0.16	0.50	1.00	0.49	-0.24	0.21
	1.00	1.10	-0.59	0.68	0.22	0.70	0.70	0.68	0.40	0.16
	0.90	1.10	-0.52	0.40	0.17	1.00	1.00	-1.46	1.02	0.17
	0.80	1.10	-1.25	-0.59	0.20	0.80	1.60	-0.59	-1.04	0.19
	0.90	1.00	-0.24	-1.04	0.20	1.20	0.80	-0.52	-1.46	0.16
	0.70	0.80	1.52	1.35	0.20	1.60	0.90	0.53	-0.17	0.20
	0.70	1.20	-0.10	1.02	0.16	0.90	0.80	-0.17	-0.66	0.19
	1.10	0.60	-0.38	0.26	0.16	1.80	0.90	0.04	-0.03	0.17
	0.60	1.40	-1.04	0.49	0.21	0.70	1.20	-1.04	0.62	0.16
	1.60	0.90	0.95	-0.24	0.21	1.30	0.80	-1.25	-0.45	0.18
	1.80	1.10	-1.46	0.95	0.20	0.80	0.90	-0.24	0.95	0.16
	1.00	1.30	1.02	0.53	0.22	1.20	1.20	0.18	1.35	0.20
	0.90	0.90	0.53	-0.66	0.22	0.90	1.20	0.74	0.26	0.21
	1.30	1.60	1.35	-0.52	0.21	1.10	0.80	-0.66	-0.59	0.16
	1.40	0.90	0.04	-0.10	0.16	0.80	0.60	-0.38	-0.38	0.20
	0.80	1.00	-0.03	0.04	0.17	1.10	0.70	0.11	0.74	0.21
	1.00	0.70	0.26	-0.31	0.17	1.40	1.30	0.40	1.52	0.20
	1.10	1.00	0.74	-0.45	0.19	1.20	1.10	1.52	0.68	0.22
2	0.80	0.00	1.02	-0.03	0.17	0.80	0.00	-0.45	-0.52	0.18
	0.80	0.00	0.40	-0.24	0.16	1.60	1.60	1.35	1.35	0.22
	1.10	0.00	-0.59	-0.59	0.21	0.70	0.00	1.02	0.26	0.16
	1.00	0.00	0.68	0.32	0.20	0.90	2.00	-1.25	-1.25	0.16
	1.20	0.00	0.11	-0.84	0.17	0.70	0.00	0.95	-0.45	0.20
	1.80	0.00	-2.00	-1.04	0.19	1.40	1.40	0.32	0.32	0.16
	1.40	0.00	-0.17	-0.66	0.22	1.00	0.00	0.11	0.18	0.18
	0.70	0.00	0.32	0.49	0.16	1.10	0.00	1.52	-0.84	0.20
	1.60	0.00	0.18	0.11	0.20	1.80	0.00	-2.00	-0.31	0.16
	1.00	0.00	-0.84	-1.46	0.20	1.00	0.00	-1.04	0.40	0.20
	1.10	0.00	-0.10	0.04	0.16	1.00	0.00	-0.24	0.62	0.19
	0.80	0.00	0.49	-2.00	0.20	1.20	0.00	-0.17	-1.46	0.21
	0.80	0.00	-0.38	-0.45	0.20	1.10	0.00	-0.38	-0.17	0.20
	1.00	0.00	-1.25	0.74	0.22	0.90	0.00	0.62	-0.03	0.20
	1.00	0.00	0.53	-0.38	0.18	1.20	0.00	0.04	-2.00	0.22
	1.10	0.00	0.95	0.95	0.16	1.10	0.00	-0.31	0.53	0.19
	0.50	0.00	0.04	-0.31	0.16	1.20	0.00	-1.46	1.52	0.17
	0.90	0.00	0.62	0.62	0.20	0.90	0.00	-0.03	-0.59	0.20
	1.20	0.00	-0.31	-1.25	0.20	0.60	0.00	0.40	-1.04	0.18
	0.60	0.00	0.26	0.18	0.19	1.00	1.80	0.49	0.49	0.21
	2.00	0.00	-0.66	0.53	0.18	1.30	1.30	-0.10	-0.10	0.1[illegible]
	1.10	0.00	-1.04	1.02	0.21	0.80	0.00	-0.52	0.04	0.17
	0.90	0.00	1.35	-0.10	0.18	0.80	0.00	0.25	-0.38	0.16
	0.70	0.00	-0.03	1.35	0.20	1.10	0.00	-0.59	0.11	0.20
	0.90	0.00	-1.46	0.40	0.17	1.00	0.00	0.18	0.74	0.20
	1.20	0.00	1.52	0.26	0.16	0.90	0.00	0.68	1.02	0.19
	0.60	0.00	-0.45	0.68	0.22	0.80	0.00	0.53	0.95	0.17
	1.30	0.00	-0.52	-0.52	0.17	1.00	0.00	-0.66	-0.66	0.16
	1.00	0.00	-0.24	1.52	0.16	0.60	0.00	-0.84	0.68	0.21
	0.90	0.00	0.74	-0.17	0.21	0.50	0.00	0.74	-0.24	0.22

(table continues)

Table 3 (continued)

Configuration	Easy Test					Hard Test				
	AL1	AL2	BL1	BL2	CL	AH1	AH2	BH1	BH2	CH
3	1.10	0.00	0.68	-0.03	0.18	0.80	0.90	0.53	-2.00	0.16
	2.00	1.60	0.11	1.02	0.19	0.90	1.10	-0.03	0.49	0.19
	1.20	0.00	0.62	-1.04	0.17	2.00	1.40	1.52	-0.45	0.16
	1.20	0.00	-0.17	0.68	0.19	1.40	1.00	1.35	-0.84	0.20
	0.50	0.00	0.26	0.62	0.16	0.60	1.60	0.68	0.95	0.18
	1.40	1.80	0.40	1.35	0.20	1.00	1.20	1.02	-1.04	0.22
	1.00	0.00	0.18	0.04	0.16	1.80	1.10	0.04	0.40	0.16
	1.00	0.00	-0.66	-1.46	0.16	1.00	0.80	-0.52	0.11	0.18
	1.10	0.00	-0.38	0.18	0.18	1.10	0.90	0.49	[illegible]	0.20
	0.90	0.00	-0.84	0.11	0.20	0.60	1.10	0.74	0.32	0.20
	0.70	0.00	-2.00	-0.24	0.20	1.10	0.70	0.32	-0.38	0.20
	0.90	0.00	-0.45	-1.25	0.20	0.80	1.00	-0.45	0.04	0.21
	0.80	0.00	0.53	-0.17	0.16	0.50	1.00	0.26	0.74	0.21
	1.20	0.00	-0.59	-2.00	0.22	1.00	0.70	0.95	0.53	0.16
	0.60	0.00	-0.52	0.32	0.17	1.20	1.00	-1.04	1.02	0.17
	0.80	0.00	-1.04	0.53	0.20	1.00	0.80	-0.59	-1.25	0.19
	1.10	0.00	-0.31	-0.31	0.20	1.20	0.50	-2.00	-0.24	0.16
	1.10	0.00	1.52	-0.84	0.20	1.60	1.80	0.62	1.52	0.19
	0.80	0.00	-0.10	-0.10	0.16	0.80	0.80	-0.66	-0.66	0.20
	0.90	0.00	-0.24	0.95	0.16	0.90	0.90	-0.84	-0.03	0.18
	0.60	0.00	-1.25	0.49	0.22	0.70	1.20	-1.46	-0.31	0.16
	1.00	0.00	0.04	-0.26	0.21	1.30	0.90	-1.25	0.18	0.17
	1.60	1.20	-1.46	-0.45	0.20	0.80	1.00	-0.24	-0.52	0.17
	1.00	0.00	1.35	-0.[illegible]	0.22	0.70	2.00	0.18	0.62	0.22
	0.90	0.00	0.49	-0.[illegible]	0.21	0.90	1.20	-0.10	0.26	0.21
	1.30	2.00	1.02	1.5[illegible]	.21	1.10	0.80	-0.17	-0.59	0.16
	1.80	1.40	0.95	0.40	.16	1.20	0.60	-0.38	-1.46	0.20
	0.70	0.00	-0.03	0.74	0.18	1.10	0.60	0.11	-0.17	0.20
	1.00	0.00	0.32	-0.38	0.17	1.00	1.30	0.40	1.35	0.22
	0.80	0.00	0.74	-0.59	0.19	0.90	1.10	-0.31	0.68	0.20
4	0.50	0.00	-2.00	-0.38	0.18	0.90	0.00	-0.45	-0.10	0.21
	0.90	0.00	-0.24	-0.66	0.20	0.90	0.00	-1.04	-2.00	0.16
	0.60	0.00	0.40	-0.10	0.20	1.10	0.00	-0.66	-1.25	0.17
	0.70	0.00	-0.59	-0.31	0.20	0.60	0.00	-0.52	-0.17	0.22
	0.70	0.00	-1.46	1.35	0.21	0.70	0.00	0.04	-0.84	0.22
	0.80	0.00	-1.04	-0.52	0.19	1.00	0.00	-0.24	-0.66	0.20
	0.80	0.00	0.53	-0.84	0.20	0.80	0.00	-0.31	-0.59	0.17
	0.80	0.00	-0.84	-0.45	0.17	0.80	0.00	-1.46	-1.04	0.20
	1.20	0.00	-0.17	-0.03	0.22	2.00	0.00	0.40	-0.31	0.20
	0.90	0.00	-1.25	0.32	0.16	1.00	0.00	1.52	0.74	0.16
	0.90	0.00	0.68	0.04	0.20	1.20	0.00	0.68	-0.45	0.22
	0.60	0.00	-0.31	-0.24	0.16	1.10	0.00	0.18	-0.24	0.18
	0.90	0.00	0.26	-0.59	0.19	0.90	0.80	-0.84	1.02	0.16
	1.00	0.00	-0.52	-1.25	0.21	1.30	0.00	-0.10	-0.52	0.16
	1.00	0.00	0.62	0.11	0.22	1.40	0.00	0.74	0.04	0.20
	1.00	0.00	0.95	0.26	0.18	0.80	0.50	-0.59	0.32	0.16
	1.00	0.00	1.52	0.62	0.17	1.80	0.70	-0.17	0.95	0.17
	1.10	0.00	-0.03	0.53	0.16	1.20	0.00	1.02	0.18	0.16
	1.10	0.00	-0.45	0.49	0.16	0.70	0.00	-2.00	-1.46	0.18
	1.00	0.00	0.49	0.68	0.17	1.10	0.00	0.11	0.32	0.20
	1.10	0.00	-0.66	0.40	0.16	1.00	0.00	0.32	0.40	0.18
	1.10	0.00	0.11	-1.46	0.20	0.60	0.00	-0.38	-0.03	0.20
	0.80	0.00	0.74	-1.04	0.16	0.80	0.00	0.53	0.11	0.20
	1.20	0.00	1.02	-2.00	0.21	0.90	0.00	0.95	0.49	0.19
	1.20	0.00	-0.10	0.95	0.22	1.60	0.00	0.26	0.53	0.20
	1.30	0.00	0.04	1.52	0.18	0.50	0.00	-1.25	-0.38	0.21
	1.40	0.00	0.32	0.18	0.20	1.10	0.00	1.35	0.26	0.19
	1.60	0.00	0.18	-0.17	0.16	1.20	0.00	-0.03	0.68	0.16
	1.80	0.00	-0.38	0.74	0.20	1.00	1.00	0.49	1.52	0.19
	2.00	0.00	1.35	1.02	0.19	1.00	0.90	0.62	1.35	0.21

(table continues)

Table 3 (continued)

Configuration	Easy Test					Hard Test				
	AL1	AL2	BL1	BL2	CL	AH1	AH2	BH1	BH2	CH
5	0.00	1.10	-0.66	0.04	0.17	0.00	1.20	-1.46	-0.66	0.18
	0.00	1.20	0.95	0.68	0.21	0.60	1.30	1.62	0.49	0.20
	0.60	1.20	-0.45	-0.31	0.17	0.60	0.00	0.40	-0.10	0.17
	0.70	0.00	-1.46	0.40	0.18	0.70	0.00	0.95	-0.45	0.20
	0.00	1.30	0.11	0.32	0.20	0.00	1.80	-2.00	1.02	0.16
	0.80	0.00	-0.31	1.02	0.16	0.80	0.00	1.52	1.52	0.20
	0.80	0.00	-0.24	1.52	0.19	0.80	0.00	0.04	-0.59	0.22
	0.80	1.10	0.40	-0.45	0.19	0.80	0.00	0.74	-1.25	0.21
	0.00	1.40	0.74	0.26	0.20	0.80	1.10	0.68	-0.31	0.16
	0.90	0.00	0.16	0.53	0.20	0.00	1.00	-0.24	-0.24	0.20
	0.90	0.90	-0.17	-0.84	0.20	0.90	0.00	0.18	0.68	0.18
	0.90	0.00	-2.00	0.11	0.19	0.00	1.10	1.35	1.35	0.20
	0.00	1.00	1.35	0.95	0.17	0.00	1.20	-1.04	-0.52	0.16
	1.00	0.00	0.49	-1.25	0.22	1.00	0.00	0.49	0.04	0.17
	0.00	1.20	0.32	-0.17	0.21	0.00	1.10	-0.66	0.18	0.16
	1.00	0.00	-0.84	-0.10	0.16	1.00	0.00	0.32	0.62	0.19
	0.00	1.00	-0.52	0.74	0.16	0.00	1.40	-0.84	0.32	0.20
	0.00	1.00	-0.10	-0.59	0.20	1.00	0.00	-0.45	0.11	0.19
	1.10	0.00	-0.59	0.18	0.20	1.10	0.90	0.26	-1.04	0.22
	1.10	0.00	-1.04	-0.66	0.16	1.10	0.00	-0.03	0.95	0.17
	0.00	0.80	0.04	-0.38	0.18	0.00	0.90	0.11	-0.03	0.16
	1.10	0.50	0.62	0.49	0.22	0.00	0.90	0.53	0.53	0.16
	0.00	0.80	1.02	-0.52	0.16	1.20	0.00	-0.59	0.26	0.19
	1.20	0.00	0.26	1.35	0.18	0.00	1.00	-0.31	-1.46	0.21
	1.20	0.00	1.52	-0.03	0.16	0.00	0.70	-0.52	0.40	0.20
	0.00	0.70	-1.25	-0.24	0.16	1.30	0.00	-0.10	0.74	0.18
	0.00	0.90	-0.03	-1.04	0.20	0.00	0.80	-0.17	-0.38	0.22
	1.60	0.00	0.53	-2.00	0.21	1.60	0.50	-0.38	-0.17	0.16
	0.00	0.60	-0.38	-1.46	0.20	0.00	0.60	1.02	-2.00	0.21
	2.00	0.00	0.68	0.62	0.22	2.00	0.00	-1.25	-0.84	0.20
6	0.50	0.60	0.32	-0.10	0.16	0.50	1.60	0.32	0.53	0.16
	0.60	0.50	0.04	-0.45	0.21	0.60	1.30	-0.84	-0.84	0.16
	0.60	0.60	-0.10	0.11	0.19	0.60	1.40	-0.66	-1.04	0.20
	0.70	0.70	-0.59	0.32	0.20	0.70	2.00	-0.52	0.26	0.18
	0.70	0.70	0.74	0.62	0.19	0.70	1.80	0.11	0.62	0.16
	0.80	0.00	0.53	-2.00	0.21	0.80	0.00	-0.59	-0.66	0.17
	0.80	0.00	0.68	-1.46	0.17	0.80	0.00	0.62	-0.45	0.21
	0.90	0.00	-0.38	-0.03	0.16	0.90	0.00	0.26	1.02	0.20
	0.80	0.00	0.62	1.35	0.16	0.80	0.00	1.35	-0.17	0.17
	1.00	0.00	-0.66	-0.84	0.16	0.90	0.00	0.40	-0.52	0.16
	0.90	0.00	0.95	-0.59	0.20	0.90	0.00	1.52	-1.46	0.21
	0.90	0.00	1.02	0.40	0.18	0.80	0.00	0.04	0.68	0.16
	0.80	0.00	0.49	-1.04	0.17	0.90	0.00	0.18	0.74	0.20
	0.90	0.00	0.40	-1.25	0.20	1.00	0.00	-0.24	1.35	0.16
	1.00	0.00	-1.46	-0.66	0.20	1.00	0.00	-0.10	0.11	0.18
	1.00	0.00	-0.31	1.02	0.22	1.00	0.00	-0.38	-0.31	0.22
	1.00	0.00	-0.24	0.95	0.16	1.00	0.00	1.02	0.49	0.22
	1.20	0.00	-0.17	0.26	0.16	1.00	0.00	0.74	0.95	0.20
	1.10	0.00	1.35	0.74	0.17	1.10	0.00	0.95	0.32	0.22
	1.10	0.00	-2.00	0.68	0.16	1.10	0.00	-0.03	-1.25	0.17
	1.10	0.00	-0.45	-0.31	0.20	1.10	0.00	-0.45	-0.59	0.19
	1.10	0.00	-0.03	0.49	0.18	1.10	0.00	0.49	-0.38	0.18
	1.00	0.00	0.26	0.53	0.22	1.20	0.00	0.53	0.18	0.20
	1.20	0.00	1.52	0.04	0.22	1.80	0.00	-0.31	-0.03	0.20
	1.20	0.00	0.18	1.52	0.19	1.20	0.00	0.68	-0.10	0.21
	1.30	0.00	-0.84	0.18	0.21	1.30	0.00	-2.00	-0.24	0.16
	1.40	0.00	-1.04	-0.24	0.20	1.40	0.00	-1.25	0.04	0.20
	1.60	0.00	-1.25	-0.38	0.18	1.60	0.00	-0.17	-2.00	0.19
	1.80	0.00	0.11	-0.17	0.20	1.20	0.00	-1.46	1.52	0.19
	2.00	0.00	-0.52	-0.52	0.20	2.00	0.00	-1.04	0.40	0.20
7	same	as	Configuration	1		same	as	Configuration	1	

(table continues)

Table 3 (continued)

Configuration	Easy Test AL1	AL2	BL1	BL2	CL	Hard Test AH1	AH2	BH1	BH2	CH
8	0.80	0.80	0.53	0.49	0.20	0.90	0.60	1.52	0.40	0.16
	1.10	0.80	0.68	0.18	0.18	1.60	1.30	0.26	-0.66	0.20
	1.80	0.80	1.02	0.53	0.21	1.20	1.00	0.62	0.74	0.16
	1.10	1.00	-0.45	0.26	0.16	0.80	0.60	-0.03	-1.25	0.17
	1.20	1.10	0.95	-1.25	0.20	1.00	2.00	-0.66	1.35	0.16
	1.00	1.00	0.40	-0.03	0.20	1.40	0.90	1.02	0.95	0.20
	1.10	0.90	-0.10	0.04	0.17	1.10	0.90	-0.17	0.62	0.20
	1.00	2.00	-1.04	-1.46	0.16	1.00	1.10	0.32	-0.10	0.16
	1.00	1.10	0.26	1.52	0.16	0.50	1.20	-1.25	-2.00	0.20
	1.40	1.30	1.35	1.35	0.18	1.20	0.90	0.04	0.32	0.22
	0.90	1.00	-2.00	0.40	0.16	1.80	0.80	0.11	0.49	0.17
	0.80	0.50	0.49	0.11	0.17	1.00	1.10	0.68	0.18	0.21
	0.70	1.80	-1.46	0.32	0.20	1.00	0.70	-0.45	0.68	0.20
	1.10	0.80	-0.59	0.68	0.17	0.90	1.00	-0.31	-0.59	0.18
	0.70	1.20	0.74	1.02	0.21	0.90	1.20	1.35	0.11	0.16
	0.60	0.90	0.62	-0.59	0.20	0.90	0.50	0.74	-0.52	0.18
	1.30	1.00	0.18	0.74	0.20	1.30	1.60	-0.84	-1.04	0.19
	1.20	1.60	0.04	-0.17	0.18	2.00	0.90	0.95	-1.46	0.17
	1.60	1.20	-0.24	-0.45	0.16	1.20	1.10	0.40	0.26	0.20
	2.00	0.90	1.52	-0.31	0.16	1.10	1.80	0.18	-0.31	0.21
	0.80	1.00	-0.52	-0.10	0.20	0.70	0.70	-0.24	-0.45	0.16
	0.90	0.70	-0.66	-1.04	0.19	0.80	0.80	-2.00	0.53	0.16
	1.20	0.70	-0.84	-0.24	0.22	0.80	1.00	-0.38	1.02	0.22
	0.80	1.20	-1.25	0.95	0.19	1.10	1.00	-0.10	-0.24	0.18
	0.60	1.40	-0.17	-0.66	0.22	0.80	1.00	-0.59	-0.38	0.19
	0.50	0.60	0.11	-2.00	0.19	0.60	0.80	0.53	-0.84	0.22
	1.00	0.90	-0.38	-0.52	0.21	0.60	1.40	-1.46	-0.17	0.21
	0.90	1.10	-0.31	-0.38	0.22	1.00	1.20	-0.52	-0.03	0.17
	1.00	0.60	-0.03	0.62	0.16	0.70	0.80	-1.04	1.52	0.20
	0.90	1.10	0.32	-0.84	0.20	1.10	1.10	0.49	0.04	0.20
9	same	as	Configuration		1	same	as	Configuration		1
10	0.50	0.80	1.52	1.35	0.21	0.50	0.80	-1.25	-2.00	0.17
	0.60	0.90	-2.00	-1.46	0.16	0.60	0.90	0.49	0.11	0.16
	0.60	0.70	-1.04	-1.04	0.16	0.60	0.70	0.04	0.49	0.18
	0.70	0.60	0.95	1.02	0.16	0.70	0.60	-0.24	-0.52	0.16
	0.70	0.90	-0.17	-0.31	0.19	0.70	0.90	-0.52	-0.24	0.20
	0.80	0.70	-0.66	-0.38	0.20	0.80	0.70	0.62	0.40	0.18
	0.80	1.00	-0.03	0.26	0.20	0.80	0.50	0.40	0.62	0.20
	0.80	0.50	0.26	-0.03	0.21	0.80	0.60	-1.46	-1.04	0.21
	0.80	0.90	-0.84	-0.59	0.21	0.80	1.00	0.11	0.04	0.17
	0.90	1.10	-0.45	-0.45	0.22	0.90	0.80	0.53	-0.31	0.20
	0.90	1.10	0.32	0.62	0.20	0.90	1.00	-1.04	-0.17	0.16
	0.90	0.60	0.68	0.49	0.20	0.90	1.20	-0.31	-1.25	0.22
	0.90	1.00	0.62	0.32	0.17	0.90	1.10	-0.17	0.74	0.19
	1.00	0.90	0.11	0.04	0.17	1.00	0.80	1.52	1.35	0.21
	1.00	0.80	0.04	0.11	0.20	1.00	1.20	-0.45	-0.66	0.16
	1.00	1.10	1.35	1.52	0.19	1.00	1.30	0.32	0.53	0.16
	1.00	0.70	0.40	-0.10	0.20	1.00	0.90	-0.03	-0.38	0.17
	1.00	1.20	-0.10	0.40	0.22	1.00	1.10	-0.38	-0.03	0.20
	1.10	1.00	-0.59	-0.84	0.18	1.10	0.90	-0.66	-0.45	0.19
	1.10	0.80	0.74	0.74	0.18	1.10	1.40	0.26	-0.59	0.22
	1.10	1.00	-1.46	-2.00	0.16	1.10	1.20	-0.59	0.26	0.22
	1.10	1.20	-0.31	-0.17	0.20	1.10	1.00	0.95	1.02	0.20
	1.20	1.10	-0.52	0.53	0.18	1.20	1.00	-0.84	-0.84	0.18
	1.20	1.00	0.53	-0.52	0.20	1.20	0.90	1.35	1.52	0.19
	1.20	1.30	0.49	0.68	0.22	1.20	1.10	0.18	-0.10	0.16
	1.30	1.60	-0.38	-1.25	0.16	1.30	1.00	1.02	0.95	0.22
	1.40	1.20	-0.24	-0.66	0.15	1.40	1.10	0.74	0.32	0.20
	1.60	1.40	-1.25	-0.24	0.17	1.60	1.60	-0.10	0.18	0.20
	1.80	2.00	1.02	0.95	0.18	1.80	2.00	0.68	0.68	0.15
	2.00	1.30	0.18	0.15	0.19	2.00	1.60	-2.00	-1.46	0.20

(table continues)

Table 3 (continued)

	Easy Test					Hard Test				
Configuration	AL1	AL2	BL1	BL2	CL	AH1	AH2	BH1	BH2	CH
U	0.80		0.68		0.18	0.90		0.62		0.16
	2.00		0.16		0.20	0.90		-0.03		0.18
	1.00		0.62		0.18	2.00		-0.31		0.17
	1.10		-0.17		0.20	1.00		1.35		0.20
	0.50		0.32		0.16	0.60		1.02		0.19
	1.20		0.40		0.20	1.00		0.95		0.20
	1.00		0.11		0.16	0.80		-0.84		0.16
	1.20		-0.66		0.16	1.00		-2.00		0.18
	1.20		-0.31		0.18	1.10		0.26		0.20
	0.90		-0.84		0.20	0.60		-0.10		0.22
	1.10		-2.00		0.19	1.10		0.32		0.22
	0.60		-0.45		0.19	1.00		-0.45		0.20
	0.80		0.49		0.16	0.50		0.49		0.21
	1.00		-0.59		0.22	0.70		0.68		0.16
	0.90		-0.52		0.17	1.00		-1.46		0.17
	0.80		-1.25		0.20	0.80		-0.59		0.19
	0.90		-0.24		0.20	1.20		-0.52		0.16
	0.70		1.52		0.20	1.60		0.53		0.20
	0.70		-0.10		0.16	0.90		-0.17		0.19
	1.10		-0.38		0.16	1.80		0.04		0.17
	0.60		-1.04		0.21	0.70		-1.04		0.16
	1.60		0.95		0.21	1.30		-1.25		0.18
	1.80		-1.46		0.20	0.80		-0.24		0.16
	1.00		1.02		0.22	1.20		0.18		0.20
	0.90		0.53		0.22	0.90		0.74		0.21
	1.30		1.35		0.21	1.10		-0.66		0.16
	1.40		0.04		0.16	0.80		-0.38		0.20
	0.80		-0.03		0.17	1.10		0.11		0.21
	1.00		0.26		0.17	1.40		0.40		0.20
	1.10		0.74		0.19	1.20		1.52		0.22

Note: AL1 and BL1 are discrimination and difficulty for Trait 1 on the easy test.
AL2 and BL2 are discrimination and difficulty for Trait 2 on the easy test.
AH1 and BH1 are discrimination and difficulty for Trait 1 on the hard test.
AH2 and BH2 are discrimination and difficulty for Trait 2 on the hard test.
Base item parameters are before ±0, ±1, or ±2 are added to the difficulty parameters.

For Configuration 1, both traits were assumed to be measured by both tests (Test 1, the easier test or level, and Test 2, the harder test or level). Zero correlation is assumed between Trait 1 and Trait 2. All correlations for difficulty and discrimination, both within and between traits, are assumed to be zero. This configuration was chosen as approximating a situation such as Language Mechanics where Trait 1 is end punctuation (i.e., period, question mark, exclamation mark), and Trait 2 is middle punctuation (i.e., comma, colon, semicolon). End punctuation is typically taught before middle punctuation. It is assumed that no correlation exists between Trait 1 (end punctuation) and Trait 2 (middle punctuation) since the ability to understand how to use periods, exclamation points, and question marks appears to be independent of the ability to understand commas, colons, and semicolons. Theoretically, a student could understand and/or master either trait without any knowledge of the other trait.

The concept of decimals is usually introduced after other, more basic, concepts (number, addition, subtraction, etc.) have been taught. Therefore. items measuring knowledge of decimals usually do not occur in the lowest or easiest levels of a series of tests designed to cover kindergarten through high school. Configuration 2 was chosen to represent such a situation where the second trait (such as decimals) is measured only by a few items on the harder test. For the harder test, these few items are assumed to have equal difficulty on both traits. The discrimination for these items is assumed to be medium on Trait 1 (other items) and high on Trait 2 (decimals). All other items include a range of discrimination values on Trait 1 and 0 discrimination on Trait 2. A low correlation between the two traits (knowledge of decimals and knowledge of other items) is assumed.

Configuration 3 involves a situation in which Trait 2 is measured on the easier test by a few items, and on the harder test by most or all of the items. For the easier test the discrimination is assumed to be high for these items on both traits, while all other items have a range of discrimination on Trait 1 and 0 discrimination on Trait 2. The harder test has a range of discrimination on both traits. This situation might occur if knowledge of fractions and fraction conversions to decimals was one trait measured on a test (Trait 2 here) and all other items were measuring Trait 1 (not fractions or conversions). For the easier test, a few items (the fraction/fraction conversion items) are assumed to have high discrimination on both traits. It is assumed that these few items are included in the test in order to measure knowledge of fractions and fraction conversions and hence should be highly discriminating on the trait that they are assumed to measure. It is assumed that they have high discrimination on Trait 1 (addition, for example) since conceivably, a test of knowledge of fractions would measure not only whether a student grasps the concept of fractions but also whether or not the student can add fractions.

Configuration 4 was chosen as a situation in which the second trait is measured only by a few items on the harder level. The correlation between traits is assumed to be moderate. Difficulty and discrimination are assumed to be moderately correlated for both tests on Trait 1 and highly correlated on Trait 2 on the harder test. The difficulty parameters are assumed to have a medium correlation on the harder test. This configuration was chosen to reflect a situation such as Mathematics Computation, where Trait 2 is ability with fractions (measured only by a few items on the harder test) and Trait 1 is ability with nonfraction items (measured at both levels).

For Configuration 5, both traits are assumed to be measured by both tests as in Configuration 1, except that both traits are not assumed to be measured by all items. Some items are assumed to be measuring only Trait 1 while having zero discrimination on Trait 2. Some items are assumed to measure Trait 2 only, having zero discrimination on Trait 1. A few items are assumed to measure both traits. Also, a medium correlation is assumed between traits, and discrimination is assumed to be negatively correlated across traits. This configuration was chosen to reflect a situation, such as Social Studies, where the two traits might be reading ability and ability to understand graphs. A typical Social Studies test usually contains some items pertaining to graphs only. A few items may require both reading ability and an understanding of graphs. Other items require reading only.

For Configuration 6, it is assumed that both traits are measured by both tests and the correlation between traits is .5. A few items are assumed to have low discrimination on Trait 1 for both tests, low discrimination on Trait 2 for Test 1, and high discrimination on Trait 2 for Test 2. All other items are assumed to have a mixture of high, medium and low discrimination on Trait 1 and zero discrimination on Trait 2. This configuration was chosen as being similar to a situation such as Mathematics Computation, where Trait 2 is ability in computation problems involving decimals and Trait 1 is ability in computing all problems not involving decimals. A few items involving decimals are assumed to have low discrimination on Trait 1 (no decimals) for both tests, low discrimination on Trait 2 (the decimal trait) for the lower level, and high discrimination on Trait 2 for the harder test. The 25 other items are assumed to not measure decimal ability so have zero discrimination on Trait 2.

Configuration 7 is the same as Configuration 1 (both traits are measured by both tests, randomly assign a, b, and c) except that a moderate correlation is assumed to exist between Traits 1 and 2. This configuration might result from a situation such as a Science test where an examinee's item responses could be the result of two traits: reading ability and knowledge of science facts.

Configuration 8 assumes that both traits are measured by both tests and that a high (.75) correlation exists between the two traits. Discrimination and difficulty are correlated .4 in Trait 1. A test that involves both mathematics and reading, such as Mathematics Concepts and Applications, might produce such a configuration.

Configuration 9 is the same as Configuration 1 (both traits are measured by both tests) except that a high correlation (.9) is assumed to exist between Traits 1 and 2. This configuration could represent a test such as Language Mechanics, where Trait 1 is reading ability and Trait 2 is punctuation ability. A high correlation between reading and punctuation is assumed since in order to understand the mechanics of language, both ability in reading and ability in punctuation must be present.

In a test that might involve both reading and vocabulary as separate traits, such as Reading Comprehension, a high correlation between traits would probably exist. Such a situation is assumed for Configuration 10, with discrimination and difficulty highly correlated across traits.

Configuration U is the unidimensional criterion. This configuration is simulated by setting $m = 1$ in the multidimensional model used to generate the data.

Data conditions. Combining the above described configurations results in 33 data conditions: 11 (item parameter/correlation configurations) x 3 ($\overline{b}_2 - \overline{b}_1$) values.

Simulated Data. The Simulated Data sets include three groups of simulees for each of the two traits: 2,000 of low ability, 2,000 of middle ability, and 2,000 of high ability. The 2,000 theta values for each of the three levels were generated using the IMSL multivariate normal random deviate generator, GGNSM (IMSL, 1979). In each case a normal distribution is assumed ($\overline{\theta}$ = -0.57, SD = 1.0 for the low ability group, $\overline{\theta}$ = 0.0, SD = 1.0 for the medium ability group, $\overline{\theta}$ = 0.57, SD = 1.0 for the high ability group). These differences in means were chosen to be similar to the differences between ability levels in published tests (CTBS/U, Levels E & F, and H & J). For the 33 data conditions, response vectors were generated for each of the three groups of observations for tests of 30 items each.

Separate sets of data were generated for parameter estimation and for cross-validation. For parameter estimation, data were generated for each of the 33 conditions described above. Thirty-three new sets were generated to be used for cross-validation purposes.

Response vectors. Using the prespecified "true" parameters (a_{i1}, a_{i2}, b_{i1}, b_{i2}, c_i, θ_{j1}, θ_{j2}) and the multidimensional model, $P_i(\theta_{j1},\theta_{j2})$ was computed for each observation. From these $P_i(\theta_{j1},\theta_{j2})$ values, (0,1) responses, u_{ijk}, were generated for each item i, simulee j, and test k, where u_{ijk} is 1 if a random number is less than or equal to $P_i(\theta_{j1},\theta_{j2})$ or 0 otherwise. The random number was generated from a uniform distribution using IMSL subroutine GGUBS (IMSL, 1979).

Responses were generated for all simulees for both tests (Test 1 and Test 2). For item parameter estimation, only responses to Test 1 were used for the low ability group and only responses to Test 2 were used for the high ability group. For the medium ability group, only responses to the anchor test were used. Responses for all simulees to both tests were used to examine the factor analyses.

Data verification

The means and standard deviations of the number-correct scores, item difficulties (p-values), and the KR-20 test reliability coefficients from the simulated tests were examined in order to verify that the simulations are realistic.

Verifying Multidimensionality

In order to determine whether the generated data accurately simulate real data, the following factor analyses were performed: principal component analysis of tetrachoric correlations and principal factor analysis of phi coefficients (McKinley & Reckase, 1982; Reckase, 1979). Both principal component analysis of tetrachoric correlations and principal factor analysis of phi coefficients were used.

The factor analyses were examined in terms of the proportion of variance accounted for by the first factor. This is based on the assumption that a set of items is unidimensional if a large amount of the variance is accounted for by the principal factor or component. For this study, a procedure similar to that suggested by Lord and Novick (1968, pp. 381-382) for evaluating unidimensionality by performing a principal axis factor analysis was used. The first four factors were extracted using estimated communalities in the diagonal. (The diagonal values in the correlation matrix are replaced by the squared multiple correlation of each variable with all other variables.) The items may be considered as arising from a unidimensional latent space if the first common factor accounts for a "large" proportion of the common variance and if all factors after the first account for much smaller and approximately equal proportions of the common variance.

Determination of whether the first two factors account for a "large" proportion of the common variance was done by comparing the data generated by the multidimensional model with that generated by the unidimensional model. The deviation of the multidimensional data from the unidimensional data was then determined by comparing the percent of variance accounted for by the first two factors in both sets of data.

Parameter Estimation

For each of the 33 data conditions, item parameters were estimated with LOGIST (Wingersky, Barton, & Lord, 1982). Item parameters for Test 1 were estimated using responses from the low and medium ability groups. Similarly, item parameters for Test 2 were estimated using responses from the medium and high ability groups. This allows combined samples of 4,000 simulees, a sample size that has been found to be adequate for obtaining very stable item parameter estimates (Yen,1983). Since separate pairs of LOGIST runs were made for each of the 33 data conditions, the result is two sets of estimated item parameters and estimated thetas per pair of tests. For each of these 33 conditions, the accuracy of the parameter estimates was examined.

Accuracy of the Parameter Estimation

A desirable characteristic of a parameter estimation procedure is the ability to obtain accurate item parameters. For use of IRT estimation procedures for the one-dimensional case, the assumption of unidimensionality is required in order to estimate the parameters for a given set of items and the examinee's trait levels (Lord & Novick, 1968, Ch. 16). Violation of this assumption has been suggested as a problem in estimation of item parameters (Loyd & Hoover, 1980; Cook & Eignor, 1981).

The data used here are known to exhibit multidimensionality. Therefore, it is informative to examine the effect of this multidimensionality on the estimation of the true parameters, a, b, c, and θ. This effect can then be considered when the estimated item parameters, $\hat{a}$, $\hat{b}$, and $\hat{c}$, are used to perform an equating or for other purposes.

Although in real life situations the real parameters are not known, one of the advantages of using simulated data is that the "real" item parameters are known; the "real" item parameters are those used to generate the data. Hence comparisons between real and estimated parameters can be made and such comparisons can be used to examine the accuracy of the estimation procedure.

Within level parameter estimation. Examining the accuracy of the parameter estimations within levels involves comparing in some way the multidimensional true (generating) parameters and the unidimensional estimates. The approach used by Yen (1984b) is the method used here. $P_i(\hat{\theta}_k) = P_i(\theta_{k1},\theta_{k2})$ if

$$\hat{c}_i = c_i, \text{ and} \tag{4}$$

$$\hat{a}_i(\hat{\theta}_k - \hat{b}_i) = a_{i1}(\theta_{k1} - b_{i1}) + a_{i2}(\theta_{k2} - b_{i2}). \tag{5}$$

The closed form relationship between the unidimensional estimated parameters and the multidimensional true parameters is approximated by finding the unidimensional parameters that minimize the sum of the squared differences between the two sides of Equation 5. Then

$$\hat{b}_i = \frac{a_{i1}b_{i1} + a_{i2}b_{i2}}{\hat{a}_i} \quad (6)$$

$$\hat{a}_i\hat{b}_i = a_{i1}b_{i1} + a_{i2}b_{i2}, \quad (7)$$

$$\hat{a}_i = a_{i1}r(\theta_1,\hat{\theta}) + a_{i2}r(\theta_2,\hat{\theta}), \text{ and} \quad (8)$$

$$\hat{\theta}_k = \frac{(\sum_i a_{i1}\hat{a}_i)\theta_{k1} + (\sum_i a_{i2}\hat{a}_i)\theta_{k2}}{\sum_i \hat{a}_i^2} \quad (9)$$

Equations 4, 6, 8, and 9 give the approximation of the relationship between the unidimensional estimated parameters and the multidimensional generating parameters.

Within Level Comparisons

The accuracy of the estimation procedure was examined by comparing both sets of true item parameters, from Test 1 and from Test 2, to the estimated item parameters, and by comparing true thetas to the estimated thetas using correlations. Estimated b_i values were compared with each of the two sets of true b_i values (b_{i1} and b_{i2}) for both traits. Similarily for a_i values. The estimated c_i values were compared with the one set of true c_i values. The estimated thetas ($\hat{\theta}$) were compared with the true thetas (θ_1 and θ_2) for the two traits.

Cross Validation

The Cross Validation data were generated in the same manner as were the Simulated Data (i.e. using the same data configurations and item parameters, but different seed numbers for the data generation). The Cross Validation data sets consist of a pooled group of three sets of observations: 2,000 of low ability, 2,000 of middle ability, and 2,000 of high ability. Response vectors and p-values were also generated as for the Simulated Data.

Using the fixed item parameters that were estimated from the original data, thetas were estimated for the Cross Validation data. In the first run, thetas were estimated for Test 1 using item responses for all three ability levels on Test 1 and not reached (NR) for Test 2. Item parameters were fixed at the values estimated in the parameter estimation runs for the original data. Similarily, in the second run, thetas were estimated for Test 2, using item responses on Test 2, NR on Test 1, and fixed item parameters. The accuracy of the item and ability parameter estimates was then examined in the same manner as for the Simulated Data.

RESULTS

True Item Parameters

All desired and attained correlations among the item parameters used for generating the two-trait data are within $\pm$.10 of the desired correlations, hence the attained correlations appear to be acceptable.

Simulated Thetas

Attained correlations between generated thetas on both traits are all within .04 of the desired correlations. Mean thetas are all within .08 of the desired mean ability for all three ability levels and the standard deviation of the generated true thetas are all within .05 of the desired standard deviations. The simulated thetas appear to be acceptable for Simulated and the Cross Validation data.

Simulated Item Responses

Table 4 contains means and standard deviations for number-correct scores for the pairs of simulated tests for Simulated and Cross Validation data. The simulated tests appear to be realistic, although for the $\bar{b}_2 - \bar{b}_1 = 2$ conditions, Tests 1 and 2 differ a great deal in difficulty. Recall that Test 1 and Test 2 are simulated to have equal difficulty when $\bar{b}_2 - \bar{b}_1 = 0$. When $\bar{b}_2 - \bar{b}_1 = 1$ or 2, Test 2 is the harder of the pair of tests. Tests 1 and 2 have very similar difficulties for all configurations when $\bar{b}_2 - \bar{b}_1 = 0$. For all configurations, Test 1 appears easiest and Test 2 hardest when $\bar{b}_2 - \bar{b}_1 = 2$. The Cross Validation data follow the same pattern.

Table 5 contains the KR-20 values for the pairs of tests for the Simulated and the Cross Validation data. The KR-20 values range from .74 to .94. With the exception of Configuration 3 and the medium+high ability group of Configuration 5, all KR-20 values decrease or else increase at most .01 as the tests go from easiest to hardest.

For Test 1, Configuration 3 has only five items that measure Trait 2. These five items have high discrimination on both traits. Test 2 measures both traits with all items. When $\bar{b}_2 - \bar{b}_1 = 0$, the KR-20 for Test 1 is .05 to .06 less than the KR-20 for Test 2. This contrasts with the .00 to .02 differences for all other configurations. Also, for $\bar{b}_2 - \bar{b}_1 = 0$, the Test 2 KR-20 is larger than any Test 1 KR-20 in Configuration 3.

Configuration 5 also breaks the trend of KR-20 decreasing with increasing test difficulty by having its next to smallest KR-20 on the easiest test for the medium+high group of simulees. The overall trend appears to be that the hardest test for each configuration ($\bar{b}_2 - \bar{b}_1 = 2$, Test 2) has the smallest KR-20.

Another overall trend is that Configurations 1 through 6 and the unidimensional configuration have ranges of KR-20 values from the upper 70s to the upper 80s and lower 90s for the low+medium group and a range of 80s to upper 80s and lower 90s for the medium+high group. However, Configurations 7 through 10 have overall higher KR-20 values. The low+medium group ranges from the lower 80s to lower 90s and the medium+high group are all in the lower 90s.

The Cross Validation KR-20s follow the same patterns as the Simulated Data. All cross validation KR-20s are within .02 of the original data KR-20s.

Table 4

Means and Standard Deviations of Number-Correct Scores for the Simulated Tests

			Simulation Data				Cross Validation Data			
			Ability Level							
			Low + Medium		Medium + High		Low + Medium		Medium + High	
Configu-ration	b_a-b_i	Test	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	2	1	24.03	6.05	*	*	24.15	5.95	*	*
	1		19.87	7.13	24.47	5.79	19.98	7.21	24.54	5.87
	0		15.53	7.10	20.53	7.06	15.60	7.08	20.52	7.11
	0	2	15.78	6.92	20.56	6.70	15.87	6.89	20.61	6.78
	1		11.76	6.04	16.26	6.93	12.01	6.15	16.44	6.94
	2		9.07	4.76	12.56	6.30	9.06	4.79	12.46	6.14
2	2	1	21.95	5.71	24.31	4.94	22.15	5.74	24.64	4.87
	1		19.22	6.30	22.32	5.64	19.26	6.23	22.25	5.72
	0		16.52	6.39	19.53	6.22	16.46	6.38	19.91	6.24
	0	2	16.51	6.15	19.66	6.05	16.46	6.21	20.02	6.07
	1		13.66	5.79	16.88	6.17	13.61	5.85	16.85	6.23
	2		11.22	5.09	14.06	5.92	11.26	5.08	14.05	5.84
3	2	1	21.53	5.94	24.40	5.01	21.43	5.97	24.34	5.07
	1		18.68	6.38	21.98	6.01	18.54	6.29	21.82	5.93
	0		15.56	6.14	18.97	6.45	15.73	6.23	18.96	6.33
	0	2	15.41	7.42	19.99	7.48	15.70	7.52	20.13	7.44
	1		12.08	6.58	16.39	7.54	11.80	6.43	15.97	7.48
	2		9.32	5.11	12.39	6.58	9.28	5.17	12.25	6.56
4	2	1	21.94	6.04	24.69	5.15	21.60	6.23	24.54	5.25
	1		19.04	6.49	22.17	6.04	19.06	6.53	22.22	5.98
	0		16.15	6.52	19.40	6.44	16.23	6.39	19.48	6.33
	0	2	15.33	6.43	18.72	6.62	15.46	6.31	18.78	6.44
	1		12.69	5.66	15.77	6.51	12.76	5.80	15.77	6.43
	2		10.55	4.85	12.98	5.82	10.46	4.84	12.96	5.81
5	2	1	22.69	5.53	25.46	4.40	22.73	5.49	25.51	4.35
	1		19.70	6.05	23.01	5.46	19.78	6.10	23.17	5.44
	0		16.84	5.98	20.27	5.94	16.67	6.17	20.19	6.05
	0	2	15.93	6.02	19.41	6.01	15.79	6.20	19.33	6.03
	1		12.99	5.54	16.32	6.20	13.13	5.60	16.40	6.11
	2		10.62	4.83	13.35	5.61	10.62	4.76	13.47	5.70
6	2	1	22.10	5.94	24.94	4.87	21.13	5.90	24.89	4.90
	1		19.32	6.45	22.54	5.72	19.11	6.45	22.58	5.84
	0		16.47	6.49	19.97	6.36	16.40	6.47	19.75	6.34
	0	2	16.25	6.36	19.76	6.36	16.23	6.42	19.66	6.29
	1		13.30	5.84	16.54	6.42	13.29	6.04	16.56	6.58
	2		10.60	5.02	13.58	5.99	10.64	5.12	13.56	5.96
7	2	1	23.20	7.08	*	*	23.53	6.96	*	*
	1		19.71	8.03	23.87	6.80	19.67	7.99	23.83	6.84
	0		15.79	8.00	20.33	7.89	15.75	7.92	20.13	7.83
	0	2	15.99	7.76	20.42	7.48	15.98	7.70	20.25	7.52
	1		12.40	6.94	16.44	7.75	12.37	6.92	16.45	7.67
	2		9.43	5.69	12.93	7.15	9.60	5.52	12.94	7.00
8	2	1	23.14	6.92	*	*	22.94	7.01	*	*
	1		19.42	7.81	23.49	6.73	19.37	7.91	23.49	6.76
	0		15.75	7.84	20.02	7.77	15.64	7.68	19.91	7.64
	0	2	15.50	8.01	19.88	8.11	15.36	7.82	19.71	8.03
	1		11.91	6.90	15.93	8.04	11.99	7.08	16.01	8.12
	2		9.42	5.64	12.49	7.17	9.25	5.56	12.53	7.22
9	2	1	23.25	7.28	*	*	23.01	7.41	*	*
	1		19.58	8.24	23.66	7.15	19.74	8.13	23.55	7.15
	0		16.03	8.30	20.30	8.12	16.04	8.18	20.36	8.08
	0	2	16.15	8.03	20.33	7.93	16.25	7.96	20.49	7.72
	1		12.45	7.23	16.53	8.06	12.51	7.16	16.42	8.00
	2		9.84	5.80	13.19	7.46	9.72	5.89	12.92	7.35
10	2	1	22.85	6.74	*	*	22.90	6.75	*	*
	1		19.55	7.45	23.22	6.59	19.38	7.46	22.90	6.71
	0		15.93	7.47	19.98	7.43	16.13	7.58	20.01	7.43
	0	2	15.88	7.40	19.93	7.44	16.05	7.54	19.92	7.44
	1		12.70	6.71	16.49	7.52	12.69	6.77	16.13	7.50
	2		10.16	5.34	13.29	6.82	10.24	5.66	13.29	6.88
U	2	1	21.90	5.99	24.60	5.05	21.94	5.94	24.68	4.96
	1		19.04	6.54	22.16	5.89	18.76	6.48	22.20	5.89
	0		16.31	6.33	19.59	6.35	16.32	6.29	19.58	6.28
	0	2	16.45	6.37	19.74	6.32	16.36	6.32	19.64	6.32
	1		13.66	6.00	16.65	6.30	13.46	6.03	16.60	6.49
	2		11.37	5.21	14.03	6.04	11.35	5.41	14.09	6.14

* Incomplete LOGIST run due to too many perfect scores.

Table 5

KR-20 Values of Number-Correct Scores for the Simulated Tests

			Simulation Data		Cross Validation Data	
			Ability Level			
			Low + Medium	Medium + High	Low + Medium	Medium + High
Configuration	b_2-b_1	Test	KR-20	KR-20	KR-20	KR-20
1	2	1	.91	*	.91	*
	1		.91	.91	.92	.91
	0		.90	.91	.90	.92
	0	2	.89	.91	.89	.91
	1		.85	.89	.85	.90
	2		.76	.86	.76	.85
2	2	1	.87	.86	.87	.86
	1		.87	.87	.87	.87
	0		.87	.87	.87	.88
	0	2	.85	.87	.86	.87
	1		.83	.86	.83	.86
	2		.78	.84	.77	.83
3	2	1	.88	.86	.88	.86
	1		.88	.88	.87	.88
	0		.85	.87	.86	.87
	0	2	.91	.93	.91	.93
	1		.88	.91	.87	.91
	2		.79	.88	.80	.87
4	2	1	.88	.88	.89	.88
	1		.88	.88	.88	.88
	0		.87	.88	.86	.88
	0	2	.86	.88	.88	.88
	1		.82	.87	.83	.87
	2		.76	.83	.76	.83
5	2	1	.86	.84	.86	.84
	1		.86	.87	.87	.87
	0		.84	.86	.85	.87
	0	2	.84	.86	.85	.86
	1		.80	.85	.81	.85
	2		.75	.81	.74	.82
6	2	1	.88	.87	.88	.87
	1		.88	.88	.88	.88
	0		.87	.88	.87	.88
	0	2	.87	.88	.87	.88
	1		.83	.87	.85	.88
	2		.77	.84	.78	.84
7	2	1	.94	*	.94	*
	1		.94	.93	.93	.94
	0		.92	.94	.92	.93
	0	2	.92	.93	.92	.93
	1		.89	.92	.85	.92
	2		.83	.90	.82	.89
8	2	1	.93	*	.93	*
	1		.93	.93	.93	.93
	0		.92	.93	.92	.93
	0	2	.92	.94	.92	.94
	1		.89	.93	.90	.93
	2		.84	.90	.83	.90
9	2	1	.93	*	.94	*
	1		.94	.94	.94	.94
	0		.93	.94	.93	.94
	0	2	.92	.94	.92	.93
	1		.90	.93	.90	.93
	2		.85	.91	.85	.90
10	2	1	.93	*	.93	*
	1		.93	.93	.92	.93
	0		.91	.93	.92	.93
	0	2	.91	.93	.92	.93
	1		.89	.92	.89	.92
	2		.83	.89	.84	.89
U	2	1	.88	.87	.88	.87
	1		.88	.88	.88	.88
	0		.86	.89	.86	.87
	0	2	.86	.87	.86	.88
	1		.83	.84	.84	.87
	2		.78	.84	.80	.85

* Incomplete LOGIST run due to too many perfect scores.

Tables 6 through 8 contain the results of the factor analyses for the Simulation Data and for the Cross Validation. Table 6 contains the correlations between the first two factors for the oblique rotations. There appears to be no pattern in the correlations that discriminates between the unidimensional criterion and the multidimensional data configurations. All correlations range between .33 and .60 for the principal components analyses (PCA), and between .53 and .76 for the factor analyses using squared multiple correlations in the diagonals (SMC).

A few patterns appear. For the PCA, all correlations for Test 2 decrease as the test gets harder ($\bar{b}_2 - \bar{b}_1$ increases from 0 to 1 to 2), and most correlations for Test 1 decrease as the test gets easier ($\bar{b}_2 - \bar{b}_1$ increases from 0 to 2). Also for PCA, the smallest correlations in each configuration are for the condition where $\bar{b}_2 - \bar{b}_1 = 2$. This is also true for most of the configurations for SMC. For both types of analyses, Configuration 4 has the highest correlations per condition and Configuration 5 the lowest.

Configuration 2 has the greatest range of correlations, with a spread of .22 points (.33 to .55) on PCA compared to .05 to .11 for all other configurations, and a spread of .13 points (.58 to .71) on SMC compared to .02 to .06 for all other comfigurations. One other pattern that emerges is that Configurations 2 and 7 have smaller correlations for Test 2 than for Test 1 on both PCA and SMC. All other configurations have overlapping correlations for Test 1 and Test 2.

Overall, the Cross Validation correlations follow the same general patterns as the Simulated Data. The Cross Validation correlations are at most $\pm.05$ from the corresponding Simulated Data.

Table 6

Correlations Between Factors for Factor Analyses

Configuration	b_2-b_1	Test	Simulation Data .PCA	Simulation Data .SMC	Cross Validation .PCA	Cross Validation .SMC
1	2	1	.52	.66	.50	.65
	1		.57	.69	.57	.70
	0		.57	.70	.58	.70
		2	.57	.72	.58	.70
	0		.59	.68	.55	.68
	1		.52	.69	.51	.65
	2		.46	.64	.43	.62
2	2	1	.54	.69	.55	.69
	1		.55	.71	.54	.71
	0		.55	.71	.55	.72
		2	.52	.71	.55	.72
	0		.48	.66	.52	.68
	1		.45	.65	.45	.66
	2		.33	.58	.33	.53
3	2	1	.51	.64	.52	.65
	1		.54	.67	.52	.66
	0		.56	.65	.52	.65
		2	.58	.65	.50	.65
	0		.58	.67	.57	.66
	1		.54	.69	.54	.66
	2		.48	.64	.47	.63
4	2	1	.59	.71	.59	.72
	1		.57	.72	.58	.72
	0		.57	.72	.56	.73
		2	.56	.73	.54	.73
	0		.59	.76	.57	.76
	1		.53	.79	.58	.76
	2		.51	.73	.48	.70
5	2	1	.49	.58	.49	.58
	1		.49	.58	.51	.60
	0		.49	.58	.51	.61
		2	.49	.59	.51	.61
	0		.50	.60	.51	.61
	1		.48	.57	.49	.60
	2		.43	.59	.46	.59
6	2	1	.54	.69	.53	.69
	1		.55	.70	.57	.71
	0		.55	.70	.57	.71
		2	.55	.70	.55	.71
	0		.55	.69	.54	.68
	1		.51	.62	.51	.69
	2		.44	.65	.42	.64
7	2	1	.57	.68	.58	.68
	1		.59	.68	.59	.68
	0		.59	.68	.57	.69
		2	.60	.70	.59	.69
	0		.57	.66	.57	.66
	1		.56	.67	.54	.65
	2		.50	.64	.50	.64
8	2	1	.58	.65	.58	.65
	1		.57	.65	.59	.67
	0		.57	.65	.57	.67
		2	.56	.67	.56	.66
	0		.60	.70	.60	.69
	1		.58	.69	.59	.70
	2		.53	.69	.52	.68
9	2	1	.58	.67	.59	.68
	1		.60	.69	.60	.70
	0		.60	.69	.60	.70
		2	.60	.67	.60	.68
	0		.59	.66	.59	.66
	1		.57	.66	.57	.67
	2		.50	.63	.50	.65
10	2	1	.57	.64	.56	.63
	1		.57	.65	.57	.65
	0		.57	.65	.57	.65
		2	.57	.65	.58	.66
	0		.57	.66	.57	.66
	1		.54	.66	.54	.66
	2		.48	.63	.51	.65
U	2	1	.50	.67	.49	.69
	1		.54	.71	.55	.71
	0		.53	.71	.55	.71
		2	.53	.70	.53	.69
	0		.53	.72	.54	.72
	1		.55	.73	.53	.73
	2		.46	.68	.49	.68

PCA means principal components analysis. SMC means factor analysis using squared multiple correlations for communalities.

Table 7 contains the first four eigenvalues from the principal components analyses. For all data sets, including the unidimensional, there appears to be a strong first factor and a much smaller second factor. In addition, Configurations 5, 6, and 10 appear to have a third small factor, which for Configuration 6, Test 2 is almost as large as the second factor. Configuration 9 appears to have the largest first factor and Configuration 5 appears to have the smallest.

Recall that the configurations are arranged such that the correlation between traits increases from .00 for Configuration 1 to 1.00 for Configuration U (.0, .3, .4, .5, .5, .5, .6, .75, .9, .9, 1.00). With the exception of Configurations 1, 5, 10, and the unidimensional criterion, the size of the first eigenvalue increases as the correlation between traits increases. Also, as the test gets harder within a configuration (i.e. moving from the first entry for a given configuration through the last), the size of the first eigenvalue decreases. The only exception is that the condition $\bar{b}_2 - \bar{b}_1 = 0$, Test 2 has the largest eigenvalue for Configuration 3, and for all other configurations, the condition $\bar{b}_2 - \bar{b}_1 = 1$, Test 1 has the largest.

The first eigenvalues of Configurations 1, 7, 8, 9, 10, and Test 2 of Configuration 3 are clearly greater than the first eigenvalues of the unidimensional criterion. The first eigenvalues of Configuration 5 are clearly smaller than the corresponding eigenvalues of the unidimensional criterion. The first eigenvalues of the remaining configurations (2, 4, 6, and Test 1, Configuration 3) are approximately equal to the first eigenvalues of the undimensional criterion.

In general, the Cross Validation data follow the same patterns as the Simulation Data.

Table 7

The First Four Eigenvalues from Principal Components Analyses (PCA)

			Simulated Data				Cross Validation			
			Eigenvalues							
Configuration	b_2-b_1	Test	1	2	3	4	1	2	3	4
1	2	1	9.43	1.60	1.03	.92	9.29	1.59	1.06	.97
	1		9.75	1.57	.97	.92	9.75	1.53	.96	.94
	0		9.27	1.57	1.00	.93	9.27	1.58	.95	.93
	0	2	9.94	1.78	.98	.90	9.79	1.61	.99	.92
	1		7.93	1.54	.99	.96	7.77	1.57	1.09	.97
	2		5.98	1.36	1.05	1.12	5.88	1.29	1.07	1.02
2	2	1	6.99	1.38	.95	.93	7.05	1.38	.97	.92
	1		7.23	1.36	.92	.91	7.15	1.35	.94	.92
	0		7.69	1.32	.93	.92	7.13	1.29	.93	.91
	0	2	6.74	1.35	1.08	.96	6.93	1.33	1.09	.97
	1		6.15	1.26	1.05	1.00	6.32	1.25	1.06	1.00
	2		5.35	1.18	1.06	1.02	5.22	1.21	1.04	1.02
3	2	1	7.09	1.33	.99	.96	7.08	1.32	1.01	.96
	1		7.37	1.39	.97	.91	7.20	1.40	.97	.92
	0		6.73	1.38	.99	.94	6.83	1.39	.98	.94
	0	2	6.83	1.35	.91	.85	6.92	1.86	.91	.82
	1		6.51	1.68	.96	.92	6.38	1.66	.97	.90
	2		6.29	[illegible]	1.02	.98	6.32	1.44	1.01	.97
4	2	1	7.44	1.21	.96	.94	7.72	1.23	.96	.91
	1		7.54	1.23	.95	.91	7.51	1.23	.95	.92
	0		7.15	1.20	.95	.92	6.93	1.22	.93	.94
	0	2	7.17	1.23	.95	.93	6.87	1.21	.95	.94
	1		6.20	1.23	.98	.97	6.17	1.23	.98	.96
	2		4.96	1.19	1.02	1.00	4.94	1.26	1.01	.99
5	2	1	6.66	[illegible]	1.15	.95	6.57	2.06	1.13	.99
	1		6.77	[illegible]	1.12	.97	6.94	2.04	1.16	.93
	0		6.37	[illegible]	1.14	.96	6.57	1.92	1.15	.94
	0	2	6.41	[illegible]	1.11	.99	6.39	1.95	1.10	.95
	1		5.80	1.89	1.14	.97	5.79	1.87	1.09	.97
	2		4.69	1.88	1.11	.99	4.82	1.69	1.14	1.00
6	2	1	7.54	1.46	1.02	.92	7.41	1.43	1.03	.95
	1		7.58	1.42	1.03	.88	7.64	1.39	1.02	.89
	0		7.37	1.37	1.00	.90	7.31	1.35	.99	.93
	0	2	7.30	1.47	1.25	.91	7.23	1.46	1.25	.91
	1		6.45	1.38	1.23	.95	6.66	1.42	1.19	.95
	2		5.42	1.29	1.20	.99	5.42	1.30	1.20	1.00
7	2	1	11.21	1.62	.95	.80	11.25	1.72	.92	.81
	1		11.43	1.72	.91	.77	11.33	1.68	.94	.78
	0		10.82	1.60	.93	.81	10.52	1.64	.94	.84
	0	2	10.36	1.94	.97	.79	10.17	1.90	.90	.79
	1		9.00	1.66	.98	.90	8.90	1.74	.95	.88
	2		7.32	1.52	.98	.96	7.07	1.50	.98	.97
8	2	1	11.01	1.97	.95	.78	11.15	1.96	.88	.64
	1		11.13	1.92	.91	.75	11.20	1.86	.93	.79
	0		10.56	1.74	.93	.80	10.29	1.77	.95	.81
	0	2	10.82	1.66	.90	.79	10.55	1.68	.91	.78
	1		9.10	1.59	.92	.86	9.29	1.58	.91	.87
	2		7.17	1.45	1.00	.94	7.17	1.45	.98	.95
9	2	1	11.74	1.76	.89	.74	11.86	1.67	.90	.76
	1		11.87	1.72	.89	.75	11.67	1.66	.90	.77
	0		11.14	1.69	.92	.79	11.06	1.69	.92	.80
	0	2	10.89	1.96	.85	.77	10.70	1.95	.90	.77
	1		9.58	1.76	.92	.88	9.43	1.76	.93	.90
	2		7.74	1.59	.97	.96	7.59	1.50	.97	.96
10	2	1	10.57	2.15	1.05	.88	10.62	2.18	1.07	.80
	1		10.59	2.14	1.05	.80	10.52	2.13	1.02	.81
	0		10.01	2.03	1.06	.97	10.09	2.01	1.05	.88
	0	2	9.90	1.97	1.12	.85	10.04	1.98	1.08	.85
	1		8.77	1.82	1.03	.94	8.61	[illegible]	1.08	.95
	2		6.96	1.65	1.05	1.00	7.06	[illegible]	1.03	.99
11	2	1	7.28	1.36	.96	.93	7.19	[illegible]	.96	.94
	1		7.49	1.29	.95	.90	7.48	[illegible]	.92	.89
	0		6.93	1.31	.93	.93	6.91	[illegible]	.95	[illegible]
	0	2	7.05	1.26	.95	.94	6.97	1.29	.95	.93
	1		6.34	1.20	.99	.96	6.52	1.20	.97	.96
	2		5.35	1.21	1.01	.99	5.57	1.25	1.01	.99

Table 8 contains the first four eigenvalues from the factor analyses using squared multiple correlations (SMC). SMC follows the same patterns as PCA. All data sets appear to have a strong first factor and a much smaller second factor. Configurations 5, 6, and 10 appear to have a third small factor, which for Configuration 6, Test 2 is almost as large as the second factor. Configuration 9 appears to have the largest first factor and Configuration 5 appears to have the smallest.

The size of the first eigenvalue generally increases as the correlation between traits increases, with the exception of Configurations 1, 5, 10 and the unidimensional criterion. The size of the first eigenvalue decreases as the test gets harder within a configuration except that condition $\bar{b}_2 - \bar{b}_1 = 1$, Test 1 has the largest for all configurations (except Configuration 3 where the condition $\bar{b}_2 - \bar{b}_1 = 1$, Test 2 is largest).

The first eigenvalues of Configurations 1, 7, 8, 9, 10, and Test 2 of Configuration 3 are clearly greater than the first eigenvalues of the unidimensional criterion. The first eigenvalues of Configuration 5 are clearly smaller than the corresponding eigenvalues of the unidimensional criterion. The first eigenvalues of the remaining configurations (2, 4, 6, and Test 1, Configuration 3) are approximately equal to the first eigenvalues of the undimensional criterion.

As with the principal components analyses, the Cross Validation data generally follow the same patterns as the Simulation Data.

Table 8

The First Four Eigenvalues from Factor Analyses using SMC

			Simulated Data				Cross Validation			
			Eigenvalues							
Configuration	b_2-b_1	Test	1	2	3	4	1	2	3	4
1	2	1	8.75	.86	.32	.18	8.61	.84	.38	.20
	1		9.08	.83	.25	.18	9.29	.79	.28	.19
	0		8.58	.83	.24	.14	8.58	.83	.22	.15
	0	2	8.24	1.04	.27	.13	8.27	1.07	.28	.15
	1		6.90	.73	.17	.10	7.06	.78	.21	.10
	2		5.21	.53	.18	.09	5.10	.46	.20	.08
2	2	1	6.21	.58	.10	.09	6.27	.58	.12	.07
	1		6.36	.55	.09	.06	6.37	.53	.11	.09
	0		6.21	.50	.10	.07	6.35	.48	.11	.07
	0	2	6.31	.53	.40	.06	6.19	.54	.32	.11
	1		5.78	.43	.23	.08	5.53	.43	.23	.09
	2		4.33	.33	.18	.09	4.39	.36	.16	.08
3	2	1	6.33	.58	.22	.07	6.37	.57	.25	.09
	1		6.40	.63	.18	.07	6.43	.64	.17	.07
	0		5.94	.58	.15	.08	6.05	.58	.13	.07
	0	2	9.16	1.13	.18	.14	9.26	1.15	.18	.08
	1		7.80	.92	.16	.11	7.67	.90	.15	.07
	2		5.52	.60	.10	.09	5.55	.61	.11	.08
4	2	1	6.48	.43	.08	.07	6.76	.45	.10	.08
	1		6.77	.43	.09	.08	6.74	.43	.10	.07
	0		6.37	.37	.10	.06	6.15	.37	.11	.08
	0	2	6.30	.40	.15	.08	6.09	.36	.14	.07
	1		5.39	.36	.12	.09	5.36	.37	.12	.09
	2		4.11	.31	.10	.07	4.09	.37	.08	.07
5	2	1	5.90	1.29	.36	.10	5.81	1.27	.35	.14
	1		6.01	1.33	.34	.14	6.18	1.26	.37	.11
	0		5.59	1.22	.32	.12	5.79	1.12	.33	.10
	0	2	5.64	1.18	.29	.10	5.83	1.16	.28	.10
	1		5.00	1.07	.28	.09	4.99	1.05	.24	.08
	2		3.85	.82	.21	.06	3.98	.84	.24	.06
6	2	1	6.78	.66	.24	.10	6.64	.64	.24	.10
	1		6.82	.63	.25	.08	6.88	.60	.24	.07
	0		6.62	.59	.19	.08	6.54	.55	.19	.09
	0	2	6.54	.69	.53	.10	6.47	.68	.53	.12
	1		5.66	.58	.45	.10	5.88	.62	.41	.08
	2		4.60	.45	.34	.10	4.61	.46	.35	.09
7	2	1	10.58	.93	.23	.10	10.44	1.04	.22	.12
	1		10.82	1.05	.20	.10	10.71	1.00	.22	.09
	0		10.18	.92	.15	.09	9.88	.95	.17	.08
	0	2	9.72	1.26	.20	.10	9.52	1.21	.17	.10
	1		8.32	.92	.13	.06	8.22	.99	.14	.06
	2		6.58	.73	.10	.07	6.33	.70	.10	.07
8	2	1	10.39	1.31	.25	.11	10.53	1.32	.18	.07
	1		10.32	1.25	.23	.08	10.59	1.19	.23	.08
	0		9.92	1.01	.21	.06	9.65	1.04	.24	.06
	0	2	10.18	.98	.16	.06	9.90	.99	.18	.06
	1		8.41	.86	.13	.06	8.60	.85	.15	.06
	2		6.41	.67	.13	.07	6.41	.66	.12	.06
9	2	1	11.14	1.10	.20	.08	11.27	1.01	.20	.06
	1		11.27	1.07	.18	.06	11.07	.99	.16	.07
	0		10.48	1.02	.16	.06	11.43	1.02	.19	.05
	0	2	10.27	1.29	.18	.05	10.07	1.28	.19	.05
	1		8.91	1.04	.15	.04	8.76	1.04	.14	.05
	2		7.02	.82	.13	.06	6.87	.71	.12	.07
10	2	1	9.95	1.49	.31	.14	10.01	1.52	.38	.09
	1		9.98	1.47	.34	.07	9.90	1.47	.29	.09
	0		9.38	1.35	.31	.09	9.46	1.33	.30	.08
	0	2	9.26	1.27	.38	.09	9.41	1.28	.33	.11
	1		8.09	1.08	.25	.08	7.93	1.09	.31	.08
	2		6.21	.87	.20	.07	6.32	.86	.19	.09
U	2	1	6.51	.54	.13	.08	6.42	.45	.10	.09
	1		6.72	.49	.12	.08	6.71	.51	.09	.07
	0		6.14	.49	.08	.06	6.12	.50	.08	.06
	0	2	6.27	.45	.11	.07	6.19	.48	.07	.05
	1		5.54	.30	.08	.07	5.72	.37	.08	.05
	2		4.51	.36	.08	.07	4.74	.40	.10	.07

Table 9 contains the percent of variance accounted for and the cumulative percent of variance accounted for by the first four eigenvalues. The percent of variance accounted for by the first eigenvalue ranges from 16 to 40 percent. The percentages for Configurations 1, 7, 8, 9, 10, and Test 2 of Configuration 3 are mostly in the thirties and upper twenties, while for Configurations 2, 3 (Test 1 only), 4, 5, 6, and U, the percentages are all in the upper teens and lower twenties. The Cross Validation values are all within ±.01 of the Simulated Data values and are not reported here.

Table 10 contains the percent of variance accounted for and the cumulative percent of variance accounted for by the first four eigenvalues for the factor analyses using SMC. For Configurations 2, 3 (Test 1 only), 4, 6, and U, the first eigenvalues account for at least 101% of the variance in most conditions. For Configurations 1, 3 (Test 2), 5, 7, 8, 9, and 10, the percent of variance accounted for is mostly between 90 and 100 with the exception of conditions where $\tilde{b}_2 - \tilde{b}_1 = 2$, Test 2 (the hardest test in each configuration). The second eigenvalue accounts for 6 to 10 percent of the variance in Configurations 2, 3 (Test 1), 4, 6, and U, and accounts for 9 to 14 percent of the variance for Configurations 1, 3 (Test 2), 7, 8, 9, and 10. Most notable is Configuration 5, for which the second eigenvalue accounts for 19-22 percent of the variance.

The percent of variance accounted for by the third and fourth eigenvalues is near zero for all configurations and all conditions. The cross validation data values are all within ±.02 of the Simulated Data values and are not reported here.

Table 9

Percent of Variance Accounted for by the First Four Eigenvalues for PCA for Simulated Data

			Eigenvalue 1		Eigenvalue 2		Eigenvalue 3		Eigenvalue 4	
Configuration	b_2-b_1	Test	% var	cum %	% var	cum %	% var	cum %	% var	cum %
1	2	1	31	31	06	37	03	40	03	43
	1		33	33	05	38	03	41	03	44
	0		31	31	05	36	03	39	03	42
	0	2	30	30	06	36	03	39	03	42
	1		26	26	05	31	03	34	03	37
	2		20	20	04	24	04	28	03	31
2	2	1	23	23	05	28	03	31	03	34
	1		24	24	05	29	03	32	03	35
	0		23	23	05	28	03	31	03	34
	0	2	23	23	04	27	04	31	03	34
	1		21	21	04	25	04	29	03	32
	2		18	18	04	22	04	26	03	29
3	2	1	24	24	04	28	03	31	03	34
	1		24	24	05	29	03	32	03	35
	0		22	22	05	27	03	30	03	33
	0	2	33	33	06	39	03	42	03	45
	1		28	28	06	34	03	37	03	40
	2		21	21	05	26	03	29	03	32
4	2	1	25	25	04	29	03	32	03	35
	1		25	25	04	29	03	32	03	35
	0		24	24	04	28	03	31	03	34
	0	2	24	24	04	28	03	31	03	34
	1		21	21	04	25	03	28	03	31
	2		17	17	04	21	03	24	03	27
5	2	1	22	22	07	29	04	33	03	36
	1		23	23	07	30	04	34	03	37
	0		21	21	07	28	04	32	03	35
	0	2	21	21	07	28	04	32	03	35
	1		19	19	07	26	04	30	03	33
	2		16	16	05	21	04	25	03	28
6	2	1	25	25	05	30	03	33	03	36
	1		25	25	05	30	03	33	03	36
	0		25	25	05	30	03	33	03	36
	0	2	24	24	05	29	04	33	03	36
	1		22	22	04	26	04	30	03	33
	2		18	18	04	22	04	26	04	30
7	2	1	37	37	06	43	03	46	03	49
	1		38	38	06	44	03	47	03	49
	0		36	36	05	41	03	44	03	47
	0	2	35	35	06	41	03	44	03	47
	1		30	30	06	36	03	39	03	42
	2		24	24	05	29	04	33	03	36
8	2	1	37	37	06	43	03	46	03	49
	1		37	37	06	43	03	46	03	49
	0		35	35	06	41	03	44	03	47
	0	2	36	36	06	42	03	45	03	48
	1		30	30	06	36	03	39	03	42
	2		24	24	05	29	03	32	03	35
9	2	1	39	39	06	45	03	48	02	50
	1		40	40	06	46	03	49	02	51
	0		37	37	06	43	03	46	02	48
	0	2	36	36	07	43	03	46	02	48
	1		32	32	06	38	03	41	03	44
	2		26	26	05	31	03	34	03	37
10	2	1	35	35	07	42	04	46	03	49
	1		35	35	07	43	03	46	03	49
	0		33	33	07	40	04	44	03	47
	0	2	33	33	06	38	04	43	03	46
	1		29	29	06	35	04	39	03	42
	2		23	23	06	29	04	33	03	36
11	2	1	24	24	05	29	03	32	03	35
	1		25	25	04	29	03	32	03	35
	0		23	23	04	27	03	30	03	33
	0	2	24	24	04	28	03	31	03	34
	1		21	21	04	25	03	28	03	31
	2		18	18	04	22	03	25	03	28

Table 10

Percent of Variance Accounted for by the First Four Eigenvalues for SMC for Simulated Data

			Eigenvalue 1		Eigenvalue 2		Eigenvalue 3		Eigenvalue 4	
Configuration	b_2-b_1	Test	% var	cum %	% var	cum %	% var	cum %	% var	cum %
1	2	1	97	97	09	106	04	110	02	112
	1		98	98	09	107	03	110	02	112
	0		99	99	09	108	03	111	02	113
	0	2	97	97	12	109	03	112	02	114
	1		101	101	11	112	03	115	01	116
	2		105	105	11	116	03	119	02	121
2	2	1	104	104	10	114	02	116	01	118
	1		105	105	09	114	01	115	01	116
	0		106	106	08	114	02	116	01	117
	0	2	103	103	09	112	05	117	01	118
	1		106	106	09	115	04	119	01	120
	2		109	109	09	118	04	122	02	124
3	2	1	102	102	09	111	04	115	01	116
	1		102	102	10	112	03	115	01	116
	0		104	104	10	114	03	117	01	118
	0	2	97	97	12	109	02	111	01	112
	1		99	99	12	111	02	113	01	114
	2		104	104	11	115	02	117	02	119
4	2	1	105	105	07	112	01	113	01	114
	1		106	106	07	113	01	114	01	115
	0		107	107	06	113	02	115	01	116
	0	2	106	106	06	112	03	115	01	116
	1		108	108	08	116	02	118	02	120
	2		113	113	08	121	03	124	02	126
5	2	1	92	92	20	112	06	118	02	120
	1		92	92	20	112	05	117	02	119
	0		94	94	20	114	05	119	02	121
	0	2	95	95	19	114	05	119	02	121
	1		96	96	21	117	05	122	02	124
	2		102	102	22	124	05	129	02	131
6	2	1	101	101	10	111	04	115	01	116
	1		102	102	09	111	04	115	01	116
	0		103	103	09	112	03	115	01	116
	0	2	98	98	10	108	08	116	02	118
	1		101	101	10	111	08	119	02	121
	2		105	105	10	115	08	123	02	125
7	2	1	97	97	09	106	02	108	01	109
	1		97	97	09	106	02	108	01	109
	0		99	99	09	107	01	109	01	110
	0	2	95	95	12	108	02	110	01	111
	1		99	99	11	110	01	111	01	112
	2		102	102	11	113	02	115	01	116
8	2	1	94	94	12	106	02	108	01	109
	1		95	95	11	107	02	109	01	110
	0		98	98	10	108	02	110	01	111
	0	2	98	98	09	107	02	109	01	110
	1		100	100	10	110	02	112	01	113
	2		103	103	11	114	02	116	01	117
9	2	1	96	96	10	106	02	108	01	108
	1		97	97	09	106	02	108	00	108
	0		98	98	09	107	01	108	01	109
	0	2	95	95	12	107	02	109	01	110
	1		98	98	11	109	02	111	00	111
	2		100	100	12	112	02	114	01	115
10	2	1	92	92	14	106	03	109	01	110
	1		93	93	14	107	03	110	01	111
	0		94	94	13	107	03	110	01	111
	0	2	94	94	13	107	04	111	01	112
	1		96	96	13	109	03	112	01	113
	2		99	99	14	113	03	116	01	117
U	2	1	104	104	09	113	02	115	01	116
	1		105	105	07	112	02	114	01	115
	0		106	106	08	114	02	116	01	117
	0	2	106	106	07	113	02	115	01	116
	1		108	108	07	115	02	117	01	118
	2		111	111	08	119	02	121	02	123

<u>Parameter Estimation.</u> Tables 11 and 12 contain comparisons of true versus estimated parameters. Table 11 contains the correlations of the true and estimated item parameters for the Simulated Data. For Configurations 2, 3 (Test 1), 4, 6, 10, and U, the correlations between difficulty on Trait 1 (b_1) and estimated difficulty ($\hat{b}$) are all .90 and above. In particular, the correlations for $\hat{b}$ and b_1 are all .98 and .99 for Configuration U. For Configurations 1, 3 (Test 2), 5, 7, 8, and 9, the correlations are all under .80 with most between .60 and .78. The most notable exception is that the correlations between $\hat{b}$ and b_1 for Configuration 5, Test 2 are .26, .36, and .34 in order as the test gets harder. These are the only correlations between estimated difficulty and difficulty on Trait 1 that are under .59.

For the correlations between estimated difficulty ($\hat{b}$) and difficulty on Trait 2 (b_2), the best correlations (.87 to .95) are for Configuration 10, the multidimensional configuration with the highest correlation (.90) between traits. All other correlations between $\hat{b}$ and b_2 are .78 or less. For Configurations 2, 3 (Test 1), 4 (Test 1), and 6, the correlations are all below .20. The correlations for Configuration 5, Test 1 are .24 to .30. For Configurations 1, 3 (Test 2), 4 (Test 2), 5 (Test 2), 7, 8, and 9, the correlations are all between .50 and .77.

The correlations between $\hat{b}$ and b_2 are all within $\pm.15$ of the correlations between $\hat{b}$ and b_1 for Configurations 1, 3 (Test 2), 7, 8, 9, and 10. For Configuration 4 (Test 2) these correlations are within .21 to .30 of each other and for Configuration 5 they are within .33 to .53. However, the largest differences for the correlations of estimated difficulty and difficulty on Trait 2 versus estimated difficulty and difficulty on Trait 1 occur for Configurations 2, 3 (Test 1), 4 (Test 1), and 6. These correlations range from .78 to .96 less for $r(\hat{b},b_2)$ than for $r(\hat{b},b_1)$.

Note the large differences and the increase in size of the correlations $r(\hat{b},b_1)$ and $r(\hat{b},b_2)$ for Configuration 5. For Test 1, the correlations of estimated difficulty with difficulty on Trait 1 are .77 or .78 and for Test 2 the correlations are .26 to .36. The reverse is true for Trait 2. Test 2 has the higher correlations (.67 to .77) and Test 1 has the lower (.24 to .30).

The correlations of $\hat{b}$ with $(a_1b_1 + a_2b_2)/a$ are mostly .98 and higher. A few are in the lower .90s with one .89.

The correlations of estimated discrimination $(\hat{a})$ with true discrimination on Trait 1 (a_1) follow some of the same patterns as the correlations between estimated and true difficulty. The highest correlations between $\hat{a}$ and a_1 are for Configurations 2, 3 (Test 1), 4, 6, and U, as was true for the correlations between $\hat{b}$ and b_1. However, the only configurations with correlations in the .90s are Configurations 3 (Test 1), 4, and U. Configurations 2 and 6 are in the .70s and .80s. Configurations 1, 3 (Test 2), 7, 8, 9, and 10 are all between .40 and .72, with one exception. Condition $\bar{b}_2 - \bar{b}_1 = 2$, Test 2, Configuration 7 is a very low .31. Configuration 5 has the lowest overall correlations on Trait 1 (.11 to .41), and with the exception of Configuration 4, Configuration U has the largest correlations.

For the correlations of estimated discrimination $(\hat{a})$ with discrimination on Trait 2 (a_2), the first 2 conditions of Test 1, Configuration 3 have the largest correlations (.80 and .83). All other correlations are .73 or less. Configurations 1, 8 (Test 2), 9, and 10 are mostly between .50 and .73. Correlations for Configuration 7 range from .44 to .55, and Configuration 3 (Test 2) correlations range from .25 to .55. Configuration 5 correlations are mostly in the .30s. All correlations for Configurations 2, 4, 6, and 8 (Test 1) are less than .31. In particular, for Configuration 6, all correlations of estimated discrimination with discrimination on Trait 2 are negative.

Note that, except for Configuration 10, the same configurations (2, 3 Test 1, 4, 6, and U) have the best correlations for Trait 1 discrimination as have the best correlations for Trait 1 difficulty.

Configuration U has the best overall correlations (.90 to .95) between $\hat{a}$ and $a_1b_1 + a_2b_2$. Correlations for Configurations 2 (Test 1), 3 (Test 1), 4, 5, and 6 (Test 1) are mostly in the upper .80s. Correlations for Configurations 1, 3 (Test 2), 7, and 9 are mostly in the .70s and lower .80s. For Configurations 8 and 10, the correlations are mostly in the .60s and .70s. Test 2 for Configurations 2 and 6 has the lowest correlations: .53 to .54 for Configuration 2, Test 2, and .46 to .55 for Configuration 6, Test 2. Note the difference in size of correlations between Test 1 and Test 2 for Configurations 2, 3, and 6.

There appears to be no pattern for the correlations between $\hat{c}$ and c. These correlations all range between -.05 to .79 with some of the poorest correlations ocurring for the unidimensional criterion.

With very few exceptions, all Cross Validation correlations are equal to the corresponding Simulated Data correlations. The exceptions are all within ±.02 of the Simulated Data correlations.

Table 11

Correlations of True and Estimated Item Parameters for Simulated Data

Conf	Test	b_2-b_1	$r(\hat{b},b_1)$	$r(\hat{b},b_2)$	$r(\hat{b},b^*)$	$r(\hat{a},a_1)$	$r(\hat{a},a_2)$	$r(\hat{a},a^*)$	$r(\widehat{ab},ab^*)$	$r(\hat{c},c)$
1	1	2	.74	.64	.99	.46	.73	.83	.96	.40
		1	.74	.65	.99	.51	.60	.77	.99	.35
		0	.74	.63	.99	.41	.63	.72	.98	.65
	2	0	.67	.71	.99	.56	.58	.82	.98	.43
		1	.63	.74	.99	.46	.54	.72	.95	.60
		2	.68	.73	.99	.54	.52	.77	.93	.64
2	1	2	.98	.09	.99	.88	*	.88	.97	.49
		1	.97	.13	.99	.85	*	.85	.98	.39
		0	.96	.14	.99	.74	*	.74	.99	.50
	2	0	.95	.12	.91	.84	-.01	.54	.91	.24
		1	.97	.13	.93	.73	.11	.54	.91	.11
		2	.97	.13	.89	.82	-.02	.53	.87	.51
3	1	2	.98	.12	.99	.87	.83	.91	.88	.14
		1	.98	.15	.98	.90	.80	.91	.97	.47
		0	.96	.18	.97	.92	.68	.85	.96	.19
	2	0	.76	.61	.99	.52	.55	.76	.99	.53
		1	.76	.62	.99	.60	.52	.79	.97	.59
		2	.74	.66	.99	.61	.25	.63	.95	.70
4	1	2	.99	-.10	.99	.92	*	.92	.93	.35
		1	.99	-.05	1.00	.95	*	.95	.99	.28
		0	1.00	-.05	.99	.95	*	.93	.99	.55
	2	0	.96	.66	.96	.91	.23	.88	.94	.21
		1	.94	.68	.99	.90	.30	.90	.94	.42
		2	.93	.72	.99	.89	.29	.89	.88	.35
5	1	2	.77	.24	.99	.38	.32	.93	.91	.43
		1	.78	.30	.98	.34	.31	.88	.97	.44
		0	.77	.27	.99	.11	.51	.83	.98	.22
	2	0	.26	.77	.97	.37	.39	.93	.97	-.05
		1	.36	.71	.98	.37	.37	.94	.90	.09
		2	.34	.67	.99	.41	.33	.95	.81	.66
6	1	2	.96	.02	.99	.80	-.13	.87	.97	.15
		1	.97	-.01	1.00	.78	-.20	.83	.99	.32
		0	.96	-.03	1.00	.84	-.31	.83	.99	.44
	2	0	.96	-.01	.97	.77	-.22	.45	.96	.39
		1	.97	.02	.98	.80	-.18	.55	.95	.49
		2	.97	-.01	.97	.80	-.25	.46	.90	.44
7	1	2	.74	.62	.99	.63	.44	.74	.96	.40
		1	.74	.62	.96	.64	.50	.79	.99	.45
		0	.59	.51	.99	.56	.44	.69	.98	.69
	2	0	.68	.71	.99	.61	.50	.81	.99	.42
		1	.70	.72	1.00	.56	.55	.80	.97	.79
		2	.69	.71	.99	.31	.53	.61	.94	.54
8	1	2	.65	.76	.99	.59	.30	.65	.97	.41
		1	.66	.74	.98	.58	.31	.64	.98	.38
		0	.64	.71	1.00	.41	.17	.42	.99	.55
	2	0	.62	.67	.95	.45	.52	.66	.99	.18
		1	.66	.66	.97	.61	.55	.79	.93	.28
		2	.69	.66	.99	.44	.56	.68	.86	.60
9	1	2	.74	.62	.98	.60	.42	.71	.96	.51
		1	.73	.62	.98	.56	.51	.74	.99	.63
		0	.74	.59	.99	.50	.50	.69	.99	.34
	2	0	.71	.67	.99	.41	.69	.80	.99	.56
		1	.70	.70	1.00	.48	.66	.82	.97	.72
		2	.69	.71	.99	.41	.50	.66	.94	.57
10	1	2	.93	.95	.98	.62	.64	.66	.96	.22
		1	.91	.93	.96	.71	.69	.73	.95	.39
		0	.90	.93	.97	.51	.53	.54	.96	.59
	2	0	.93	.91	.98	.67	.68	.72	.98	.67
		1	.93	.87	.94	.56	.53	.60	.96	.40
		2	.94	.94	.98	.44	.64	.57	.92	.26
U	1	2	.99	-	.99	.95	-	.95	.95	.06
		1	.99	-	.99	.90	-	.90	.99	.25
		0	.98	-	.99	.94	-	.94	.99	.40
	2	0	.98	-	.98	.90	-	.90	.97	-.13
		1	.99	-	.99	.92	-	.92	.94	.37
		2	.99	-	.94	.90	-	.90	.87	.36

Note: $b^* = (a_1b_1 + a_2b_2)/\hat{a}$. $a^* = a_1 r(\hat{\theta},\theta_1) + a_2 r(\hat{\theta},\theta_2)$. $ab^* = a_1b_1 + a_2b_2$.

Note: * Discriminations of Trait 2 are all zero.

Table 12 contains the correlations between the true and estimated trait values for both Simulated Data and Cross Validation. The correlations between estimated ability ($\hat{\theta}$) and ability on Trait 1 (θ_1) range from .44 to .92. Configurations 2, 3 (Test 1), 4, 6, and U have correlations ranging from .79 to .92 with most correlations in the .90s. For each of these configurations, the smallest correlation is for condition $\bar{b}_2 - \bar{b}_1 = 2$, the hardest test. The correlations for Configurations 3 (Test 2), 5, 7, 8, 9, and 10 are mostly in the .70s and .80s, except the hardest test of Configuration 5 with .67, and the easiest test of each of Configurations 7, 8, 9, and 10 which are .52, .55, .56, and .59, respectively. Overall, Configuration 1 has the smallest correlations between $\hat{\theta}$ and θ_1 of all the configurations, ranging from a low of .44 for the easiest test and from .62 to .68 for the other conditions.

In general, the correlations for Configurations 1, 3 (Test 2), 5, 7, 8, 9, and 10 increase as the correlation between traits increases.

The correlations of estimated ability with ability on Trait 2 are all approximately equal to the correlations of estimated ability with ability on Trait 1 for corresponding conditions of Configurations 1, 3 (Test 2), 5, 7, 8, 9, and 10. For Configurations 2, 3 (Test 1), 4, and 6, the Trait 2 correlations are all .28 to .56 less than the corresponding Trait 1 correlations. In general, these differences decrease as the correlations between traits increases.

The correlations between $\hat{\theta}$ and θ^* are mostly in the .80s and .90s. The smallest correlations are .58 for the easiest test of Configurations 7, 8, 9, and .60 for the easiest test of Configurations 1 and 10. All other correlations are .76 and above. The smallest correlations for each of Configurations 2, 3, 4, 5, 6, and U are all on the hardest test and range from .77 to .83.

For Configurations 1, 3 (Test 2), 5, 7, 8, 9, and 10, the correlations between $\hat{\theta}$ and θ^* are larger than the corresponding correlations of $\hat{\theta}$ with both θ_1 and θ_2. This difference decreases as the correlation between θ_1 and θ_2 increases. For Configurations 2, 3 (Test 1), 4, and 6, the correlations between $\hat{\theta}$ and θ^* are approximately equal to those between $\hat{\theta}$ and θ_1 and hence the correlations between $\hat{\theta}$ and θ^* are much larger than the corresponding correlations between $\hat{\theta}$ and θ_2 (since the correlations between $\hat{\theta}$ and θ_1 are much larger than the corresponding correlations between θ and θ_2 for these configurations). This difference decreases as the correlation between θ_1 and θ_2 increases.

The Cross Validation correlations are all $\pm.03$ of the corresponding Simulated Data correlations with one exception. The correlations for Configuration 2, Test 1, $\bar{b}_2 - \bar{b}_1 = 0$ is .17 points larger for the Simulated Data than for the Cross Validation data for both $r(\hat{\theta},\theta_1)$ and $r(\hat{\theta},\theta_2)$.

Table 12

Correlations of True and Estimated Traits

			Simulated Data				Cross Validation			
Conf	Test	b_2-b_1	N*	$r(\hat{\theta},\theta_1)$	$r(\hat{\theta},\theta_2)$	$r(\hat{\theta},\theta^*)$	N*	$r(\hat{\theta},\theta_1)$	$r(\hat{\theta},\theta_2)$	$r(\hat{\theta},\theta^*)$
1	1	2	4850	.44	.46	.60	4830	.47	.48	.63
		1	5277	.67	.65	.93	5216	.67	.68	.94
		0	5711	.67	.68	.91	5688	.67	.67	.91
	2	0	5703	.68	.67	.91	5712	.68	.66	.91
		1	5921	.66	.61	.85	5901	.67	.61	.84
		2	5983	.62	.55	.76	5985	.60	.52	.74
2	1	2	5491	.91	.35	.91	5473	.92	.35	.92
		1	5784	.92	.39	.92	5736	.92	.38	.92
		0	5920	.90	.36	.90	5900	.73	.33	.73
	2	0	5922	.90	.46	.91	5906	.90	.49	.91
		1	5981	.86	.49	.86	5982	.87	.46	.87
		2	5995	.82	.42	.83	5992	.81	.43	.82
3	1	2	5469	.90	.58	.92	5487	.90	.57	.92
		1	5720	.91	.57	.92	5739	.91	.56	.91
		0	5929	.90	.54	.89	5922	.89	.54	.89
	2	0	5573	.80	.76	.92	5607	.89	.77	.93
		1	5847	.76	.71	.86	5859	.77	.70	.86
		2	5965	.71	.62	.77	5970	.70	.62	.76
4	1	2	5313	.92	.49	.92	5358	.92	.52	.92
		1	5478	.92	.51	.92	5672	.92	.52	.92
		0	5887	.90	.52	.90	5870	.90	.51	.90
	2	0	5482	.90	.55	.90	5891	.90	.55	.90
		1	5949	.85	.53	.85	5959	.86	.53	.86
		2	5989	.79	.47	.79	5990	.79	.48	.79
5	1	2	5306	.78	.78	.90	5317	.79	.79	.91
		1	5680	.80	.81	.92	5685	.80	.80	.91
		0	5881	.78	.79	.89	5884	.80	.79	.90
	2	0	5932	.80	.79	.90	5953	.82	.80	.91
		1	5976	.75	.77	.85	5985	.75	.76	.85
		2	5993	.67	.70	.77	5991	.67	.70	.77
6	1	2	5416	.91	.54	.91	5387	.92	.53	.92
		1	5681	.91	.54	.91	5701	.91	.57	.92
		0	5900	.90	.56	.90	5872	.90	.54	.90
	2	0	5862	.90	.62	.91	5898	.90	.62	.91
		1	5955	.85	.58	.85	5962	.85	.58	.86
		2	5991	.81	.53	.81	5996	.81	.51	.80
7	1	2	4588	.52	.53	.58	4588	.56	.55	.61
		1	4996	.83	.82	.93	5052	.82	.82	.92
		0	5529	.81	.81	.90	5564	.81	.81	.90
	2	0	5538	.81	.81	.91	5533	.81	.80	.90
		1	5833	.76	.74	.83	5836	.76	.74	.83
		2	5945	.71	.67	.76	5951	.71	.68	.77
8	1	2	4664	.55	.55	.58	4693	.57	.56	.60
		1	5052	.86	.86	.93	5058	.86	.86	.9[illegible]
		0	5607	.81	.83	.88	5599	.81	.83	.88
	2	0	5473	.82	.85	.89	5498	.82	.84	.89
		1	5800	.80	.83	.86	5786	.80	.83	.86
		2	5932	.71	.75	.77	5931	.72	.76	.78
9	1	2	4558	.56	.56	.58	4649	.56	.56	.57
		1	4915	.90	.90	.93	4948	.90	.90	.93
		0	5448	.88	.88	.90	5464	.89	.89	.91
	2	0	5409	.88	.88	.90	5464	.89	.89	.91
		1	5755	.82	.81	.84	5788	.83	.82	.84
		2	5915	.76	.75	.77	5941	.75	.75	.77
10	1	2	4921	.59	.59	.60	5011	.60	.60	.62
		1	5261	.91	.91	.94	5321	.91	.91	.94
		0	5680	.88	.88	.91	5689	.89	.88	.91
	2	0	5646	.88	.89	.91	5666	.88	.88	.91
		1	5869	.88	.69	.90	5869	.88	.8[illegible]	.90
		2	5958	.83	.83	.85	5952	.83	.83	.85
U	1	-2	5431	.92	-	.92	5431	.92	-	.92
		1	5751	.92	-	.92	5736	.92	-	.92
		0	5595	.91	-	.91	5393	.91	-	.91
	2	0	5337	.89	-	.89	5900	.90	-	.90
		1	5121	.84	-	.84	5968	.84	-	.84
		2	5992	.80	-	.80	5963	.81	-	.81

Note: $\theta^* = [r(a_1,\hat{a})\theta_1 + r(a_2,\hat{a})\theta_2] / \hat{a}^2$.

Note: N* is total N minus perfect and/or examinees with no correct answer.

DISCUSSION

The purpose of this research is to examine the effects of various multidimensional data configurations on parameter estimation with the three-parameter logistic model. Ten two-trait data configurations and one unidimensional criterion were chosen. For each of these eleven configurations, three difficulty conditions were simulated. Data were generated using a multidimensional model for degrees of correlation between traits of .00 to .90 and one unidimensional criterion.

<u>Simulations.</u> The simulated item parameters and thetas were well within acceptable limits of the desired values. Means, standard deviations, and KR-20 values of number-correct scores indicated that all the conditions simulated realistic test configurations.

<u>Multidimensionality</u>. Two factor analyses were examined in order to verify the multidimensionality of the generated data. The factor analyses do not seem to consistently discriminate since for all data sets, including the unidimensional criterion, there appears to be a strong first factor and a much smaller second factor. Hence, all the data sets appear to be two-dimensional. The correlations between the first two factors do not appear to have any pattern that discriminates between the unidimensional criterion and the multidimensional data configurations. These correlations certainly fail to follow the pattern of correlations between traits ranging from .00 to .90 for the multidimensional configurations.

This is similar to the McKinley and Reckase (1984) findings that correlations between factors did not follow the pattern of correlations between traits for two-dimensional simulated data. However, for the McKinley and Reckase data, the size of the first eigenvalue decreased and the size of the second eigenvalue increased as the correlation between the two traits decreased. This is in direct contrast to the general trend that can be seen in the data presented here. In general, as the correlations between traits decreases, the size of the first eigenvalue increases. The multidimensional model used by McKinley and Reckase to generate their data is an extension of the Birnbaum (1968) two-parameter model that uses two discrimination parameters and one item parameter related to difficulty. Clearly, the multidimensional model used here and the multidimensional model used by McKinley and Reckase are generating different data configurations.

Upon examining the first four eigenvalues of the principal components analyses (PCA) and the factor analyses using squared multiple correlations (SMC), a few patterns emerged that caused a rethinking/restructuring of the multidimensionality (or lack of it) for each of the chosen configurations. The 'true' item discriminations for each test were chosen to represent real data in that if an item does not measure a trait on a test then the discrimination for that item on that trait is zero. If an item does measure a trait on a test, then the discrimination of that item on that trait is non-zero.

The item discriminations for each test for each configuration were examined from the point of view that if most or all of the items "load" (discriminate) on one dimension only, then the test is probably unidimensional or near enough so to be called unidimensional. If most or all of the items "load" on both dimensions, then the test is probably multidimensional.

This causes a grouping of the multidimensional configurations into three groups (actually four), based on the dimensionality of the tests. Clearly, both tests of Configurations 1, 5, 7-10, and Test 2 Configuration 3, are multidimensional since for Configurations 1, and 7-10, every item measures both dimensions, and for Configuration 5, 17 items (over half) measure Trait 1 and 17 items measure Trait 2. Hence, Group M1, a multidimensional group, consists of Configurations 1, and 7-10. Group M2, also a multidimensional group, consists of Configuration 5 alone since it is the only configuration with about half of the items measuring each dimension.

Another group (Group U) can be considered as a group with unidimensional tests. Group U consists of Configurations 2, 4, 6, and U. Both tests on these configurations and Test 1 Configuration 3 can be expected to be unidimensional because Trait 2 is either not measured at all or is measured by only five of the 30 items, (i.e., Discriminations on Trait 2 are all zero or only five items "load" on Trait 2). For Test 1 of Configurations 2 and 4, Trait 2 is not measured at all, hence has zero discriminations for all items. Therefore, Test 1 for Configurations 2 and 4 are expected to be unidimensional. Similarily, Test 1 of Configurations 3 and 6, and Test 2 of Configurations 2, 4, and 6 are all probably unidimensional since 25 of 30 items do not load on the second trait.

The two tests of Configuration 3 fit in different groups (Test 1 is unidimensional and Test 2 is multidimensional). For the sake of brevity of discussion, Test 1 will be considered as part of Group U and Test 2 as part of Group M1, although, strictly speaking, the groups consist of two-dimensional configurations, not individual tests.

The factor analyses were then reexamined taking these groupings into account. The first eigenvalues of Group M1 are all greater than the first eigenvalues of the unidimensional criterion. For Group M2, the first eigenvalue is clearly less than the first eigenvalue of the unidimensional criterion. For Group U, the first eigenvalue is approximately equal to the first eigenvalue of the unidimensional criterion. Recall that the tests of Group U are considered unidimensional by the discrimination (loading) criterion. Note, then, that these factor analysis results do support the grouping of the configurations into those with multidimensional tests versus those with unidimensional tests. When the configurations consist of unidimensional tests, the first eigenvalues are approximately equal to the first eigenvalue of the unidimensional criterion. When the configurations consist of multidimensional tests, the first eigenvalues are either larger than or smaller than the first eigenvalues of the unidimensional criterion.

Configuration 5 is similar to the McKinley and Reckase Test 1 data and the Group M1 configurations are similar to the McKinley and Reckase Test 2 data. In particular, keeping in mind that different generating models are involved, the McKinley and Reckase Dataset 2 with a correlation of .5 between traits is similar to Configuration 5 which also has a correlation of .5, and McKinley and Reckase Dataset 8 has the same correlation (0) as Configuration 1. For Dataset 2, the correlation between factors for the PCA was -.59, compared to correlations of .45 to .50 for Configuration 5. The correlations for Configurations 8 and 1 versus Datasets 5 and 8 are .55 to .60 and .46 to .57 versus .62 and -.57. These latter correspond more closely than the Configuration 5 versus Dataset 2 correlations. The McKinley and Reckase Test 2 correlations between factors varied as correlations between traits decreased, as is also true for the corresponding data sets here (Group M1 configurations).

The first four eigenvalues for Dataset 2 were 9.09, 1.79, 1.30, and 1.28, indicating a strong first factor and a smaller second one. In contrast, the Configuration 5 eigenvalues indicate three factors, a strong first factor and weaker second and third factors. Again, the generating models seem to be simulating different things. For all McKinley and Reckase Test 2 datasets, the first four eigenvalues indicated a large first factor and an extremely small or nonexistent second factor. The size of the first eigenvalue decreases and the size of the second increases as the correlation between traits decreases. Although not as clear, the same general trend of a decrease in the first eigenvalue as correlations between ability decreases appears in the data reported here for Group M1. However, Group M1 appears to have a small second factor, while the McKinley and Reckase Test 2 data do not.

The percent of variance accounted for by the first eigenvalue for the PCA also supports the pattern of groupings. Group M1 percentages are mostly in the 30s and upper 20s. Group U and the unidimensional criterion percentages are all in the tens and lower 20s. Group M2 also has percentages in the tens and lower 20s.

For the SMC, the percent of variance accounted for discriminates between these groups even better than the PCA does. For Group U and the unidimensional criterion, the first eigenvalues accounted for at least 101% of the variance in most conditions, and the second eigenvalue accounts for 6-10% of the variance. For Group M1, the percent of variance accounted for by the first eigenvalue is between 90 and 100, except for the hardest test in each configuration, and the second eigenvalue accounts for 9-14% of the variance. The first eigenvalue for Group M2 accounts for 92-102% of the variance. The second eigenvalue for Group M2 accounts for 19-22% of the variance.

Therefore, clearly, Configuration 5 (Group M2) is multidimensional, with a dominant first factor. The multidimensionality of Group M1 is also verified by the factor analyses using SMC. The SMC analyses support unidimensionality for Group U and the unidimensional criterion. (I.e. Those multidimensional configurations that appear to have unidimensional tests appear to be as unidimensional as the unidimensional criterion. Those multidimensional configurations with multidimensional tests are supported by the SMC factor analyses as being composed of two-dimensional tests as they were constructed to be.)

The multidimensionality of Configuration 5 supports the McKinley and Reckase (1984) conclusions that when the two dimensions underlying the tests are independent of each other (i.e. each item discriminates on only one of the dimensions) then correlated abilities tend to yield response date with a dominant component.

McKinley and Reckase also found that when the two dimensions underlying the test do not operate independently of each other (each item discriminates on both dimensions), then the effect of the correlations between abilities is the same, but less extreme, (i.e. correlated abilities tend to yield response data with a dominant component). The Group M1 data also appear to yield a dominant conponent. However, in contrast to the McKinley and Reckase Test 2 data having extremely small or no second factors, the Group M1 data appear to have a small second factor and in some cases a third factor.

It appears that the size of the correlation between traits used in generating the data with the multidimensional model used here was not as important in causing the data to be multidimensional as was the pattern of the loadings of the discriminations on the traits for the two tests. However, with the McKinley and Reckase model, the dimensionality of the response data appears to depend on the ability correlations. Correlated abilities tended to yield response data with correlated dimensions (tended to be unidimensional) and uncorrelated abilities tended to yield response data with relatively uncorrelated dimensions (tended to be multidimensional).

Parameter estimation. How well the item parameters were estimated appeared to depend to some extent on whether or not the tests were unidimensional or multidimensional. For Group U and the unidimensional criterion, b_1 was estimated well. Configuration 10 also has very good estimates for b_1. For Group M1, b_1 was not estimated as well. Especially notable is the extremely poor estimation of b_1 for Configuration 5, Test 2 (Group M2).

Configuration 10 had the best estimation of b_2. This would be expected since the correlation between b_2 and b_1 is .80. The poorest estimations of b_2 occurred for Group U (except Configuration 4 Test 2), and for Configuration 5, Test 1. Group M1 (except Configuration 10), Configuration 4 Test 2, and Configuration 5 Test 2 estimates, while also poor, were a little better than Group U.

Since LOGIST produces only one b, then b has to estimate both b_1 and b_2. If the correlation between b_1 and b_2 is low, either b can estimate b_1 well, or b can estimate b_2 well, or it can estimate both poorly, but it cannot estimate both well. If the correlation between b_1 and b_2 is medium, then the estimate of both b_1 and b_2 by b can be medium to poor, or possibly one can be estimated well and the other medium to poor. If the correlation between b_1 and b_2 is high, then the estimation of both b_1 and b_2 must be about the same, ranging from poor to good.

If the correlation between b_1 and b_2 is high, then LOGIST is trying to estimate practically the same difficulty values. If LOGIST did not estimate both b_1 and

b_2 well, then perhaps LOGIST would be suspected of doing a poor job of estimating the difficulty parameter.

Note in Table 1 that the correlation between b_1 and b_2 is zero for all configurations except Configuration 10, where the correlation is .8 and Configuration 4 Test 2, where the correlation is .5. For Configuration 10, since the correlation between b_1 and b_2 is so high, the excellent estimation of both would be expected. Similarily, the medium correlation between b_1 and b for Configuration 4 Test 2 could be expected since the correlation between b_1 and b_2 is .5. However, since there is zero correlation between b_1 and b_2 for all other conditions, and b_1 was well estimated for Group U, then b_2 could not be estimated very well.

Both b_1 and b_2 were estimated very poorly for Configuration 5. Note the large differences and the increase in size of the correlations $r(b,b_1)$ and $r(b,b_2)$. For Test 1, the correlations of estimated difficulty with difficulty on Trait 1 are .77 or .78 and for Test 2 the correlations are .26 to .36. The reverse is true for Trait 2. Test 2 has the higher correlations (.67 to .77) and Test 1 has the lower (.24 to .30). These two sets of correlations are the lowest of all configurations for the corresponding traits. Configuration 5 Test 2 has the lowest correlations for Trait 1 and Configuration 5 Test 1 has the lowest for Trait 2 for the multidimensional tests. For Group M1, the estimation of b_2 was mediocre, not as poor as the Group U estimates and nearly as good as the Group M1 estimates.

In spite of the fact that LOGIST is intended for only unidimensional tests, it has done an excellent job of estimating difficulty for these various multidimensional data sets. When the correlation between b_1 and b_2 was high, the estimation of both b_1 and b_2 was good. When the correlation between b_1 and b_2 was medium, b_1 was well estimated and b_2 was estimated neither poorly nor well. Similarly, when the correlation between b_1 and b_2 was zero, and Trait 2 was measured by few or no items, then the b parameter was estimated well for the trait that was measured and very poorly for the trait that was not measured. This would clearly indicate support for the belief that LOGIST is doing the task for which it is intended, at least, as far as estimating difficulty is concerned.

The differences between the correlations $r(\hat{b},b_1)$ and $r(\hat{b},b_2)$ also tend to follow the grouping pattern. The largest differences are for Group U, where b_1 is estimated very well and b_2 is estimated extremely poorly. For Group M2, Test 1 Configuration 5 has fairly good estimates of b_1 and poor estimates of b_2, while Test 2 has fairly good estimates of b_2 and poor estimates of b_1. This follows logically from the fact that nearly half the items (13 of 30) do not measure Trait 1, 13 others do not measure Trait 2, and only 4 items measure both traits. Therefore, approximately half the items "load" (discriminate) on Trait 1 and half on Trait 2. This allows one trait to be estimated fairly well while the other is estimated poorly.

It appears that how well the difficulty parameter is estimated depends on an interaction between whether or not the test is unidimensional according to the "loading" criterion, and how closely correlated the difficulty between the two traits is. If the item clearly measures one trait and not the other, the difficulty parameter on the trait measured is estimated well. However, if the test measures both traits, then mediocre estimation of the difficulty parameter for both traits can be expected unless the correlation between difficulty on both traits is high.

In all cases, $r(\hat{b},b^*)$ is very high. This supports the accuracy of Yen's (1984b) equation for predicting how the unidimensional difficulty parameter is related to the multidimensional difficulty parameters.

The estimation of the discrimination values follow some of the grouping patterns established. Discrimination for the unidimensional criterion was estimated better than for any other configurations. For Group U, a_1 is estimated well. For Group M1, the estimation of a_1 is mostly mediocre. Group M2 has the poorest overall correlations for discrimination on Trait 1.

The estimation of discrimination of Trait 2 (a_2) does not seem to follow a useful pattern. Most correlations are mediocre to poor. In general, the estimation of a_2 for Group M1 is better than that for Group U. The size of the correlation between a_1 and a_2 appears to have no effect on the estimation of either a_1 or a_2. However, the estimation of a_1 and a_2 does appear to be dependent on whether or not the tests are considered to be unidimensional according to the discrimination "loading" criterion. If the test is considered to be unidimensional according to this criterion, then a_1 is well estimated and a_2 is poorly estimated. If the test is considered to be multidimensional, then the estimation of both a_1 and a_2 is mediocre to poor.

The instability of the discrimination parameter shown here is comparable to Yen's (1980) findings of unstable item discrimination estimates found for items from an achievement test. Yen hypothesizes that a possible cause for the instability in the estimations for the real data could be a carefulness dimension. However, Yen used very small sample sizes (183-668), which have been shown to yield unstable parameter estimates.

The results found here for the discrimination estimates for Group U configurations support the Reckase (1977, 1979) conclusion that the three-parameter model computes item discrimination parameter estimates related to one factor. This is the result that would be theoretically predicted (Christoffersson, 1975). Results for the other multidimensional configurations are not so clear cut. If LOGIST were drawn to one dominant factor, one would expect the discriminations for one of the two traits to be estimated better than the other. However, in the M1 and M2 groups, this is not the case. In all conditions, the discrimination for both traits showed, at most, mediocre estimation. This supports the Drasgow and Parson's (1983) conclusion that for some multidimensional data configurations LOGIST is not drawn to a general factor.

Equation 15 predicts that if both traits are equally influencial, then the discrimination of both traits will be given equal weight in obtaining $\hat{a}$. The low correlations between $\hat{a}$ and both a_1 and a_2 for Group M2 support the accuracy of this equation. However, $r(\hat{a},a^*)$ is very high for all conditions of Group M2. As with the difficutly parameters, the accuracy of Yen's (1984b) equations for predicting the relationship between the unidimensional and the multidimensional item parameters is upheld.

Note also that it is assumed here that a high correlation implies that the parameters are well estimated. This might not be true. If the two sets of discriminations had equal standard deviations but different means, the correlation could be 1.00, even though none of the corresponding values are equal. However, this would not change the conclusions drawn from the low correlations found here. Obviously, if LOGIST were drawn to one group factor (i.e. computes item discrimination estimates related to one factor), then the item discrimination parameters for one of the two traits should have been better estimated.

The estimation of the c parameter is poor to mediocre for all conditions of all configurations. This is consistent with the Ree and Jensen (1983) results of low correlations between c and estimated c (.031 to .315) for sample sizes from 250 to 2000. Their data were generated using the one-dimensional three-parameter model. No comparisons could be made for c estimated from data generated with a three-parameter multidimensional model, since none of the multidimensional research reported data for c.

In order to adequately estimate the guessing parameter a substantial number of low ability examinees are required (Lord, 1975; Hambleton & Martois, 1983; Ree & Jensen, 1983; Wingersky, 1983). For very easy items or items that do not discriminate well, the item response function will not become asymptotic at the lower end of the range of abilities in the sample. If there are no or few examinees at the lower end of the range of abilities, there is no information with which to estimate c. Hence, LOGIST estimates a (the same) fixed c for all such items. There may have been too few examinees with low test scores for this data. However, this appears implausible as explained below.

It would be expected that as the difficulty of the items increases there would be more examinees with low test scores. Hence, c should be better estimated for the harder tests. This certainly does not appear to be the case here. Recall that the difference in average difficulty of the easiest to the hardest tests in each configuration is 2.00. The difficulty parameter for each item had 1.0 added or subtracted to obtain Test 1 and Test 2 with a 2.0 difference between average difficulty. Therefore, every item for the harder test of the each configuration has a difficulty value that is 2.0 larger than the difficulty value for some item on the easier test. Hence, c should be better estimated for the harder tests. The correlations between c and estimated c do not increase as the test gets harder. In fact for some configurations, the opposite occurs. In general, no consistent pattern occurs at all.

Lord (1975) and others have shown that LOGIST parameter estimates for the three-parameter model are adequate if $N \geq 1000$ and $n \geq 50$. The item parameters used here were estimated using 4000 simulees (2000 low + 2000 medium). (All 6000 simulees were not used since the program used to get the response vector data from tape to LOGIST could not handle 6000 simulees and 60 items.) Possibly the 30-item tests were too small for good parameter estimation.

The correlations between true and estimated trait values also follow the multidimensional/unidimensional grouping. Group U and the unidimensional criterion have the highest correlations between estimated ability and ability on Trait 1. Groups M1 and M2 have correlations about .1 to .2 lower. Configuration 1 has the lowest. Notice for Groups M1 and M2, that in general, the correlation between estimated θ and θ_1 increases as the correlation between θ_1 and θ_2 increases.

The difference between Groups M1 and M2 versus Group U is clear cut when the correlations between estimated θ and θ_2 are compared to the correlations between estimated θ and θ_1. For Groups M1 and M2, these correlations are nearly equal; for Group U, the correlations of estimated ability with ability on Trait 2 are all .28 to .56 less than the corresponding correlations of estimated ability with ability on Trait 1. These differences mostly decrease as the correlation between traits decreases.

Apparently, when both tests are unidimensional according to the "loading" criterion, ability on Trait 1 is well estimated (as well as the unidimensional criterion is estimated), but Trait 2 is very poorly estimated. These differences decrease (i.e. Trait 2 is estimated better) as the correlation between traits increases. When one or both tests are considered to be multidimensional, then the estimation of both traits is approximately the same. Also, when one or both tests are considered to be multidimensional, as the correlations between traits increases, the correlations between estimated θ and true θ increase until they are as good or better than the correlations for the corresponding configurations of the unidimensional criterion (except for the easiest test where the correlation remains less than .60).

For most tests, the correlation decreases as the test gets harder. The reason for the ocurrence of the noticeably smaller correlation on the easiest test in each of the Group M1 configurations appears to be due to excessive loss of simulees due to zero or perfect scores. Examination of Table 12 shows that five configurations (1, 7, 8, 9, 10) have a correlation under .60 for the easiest test for both traits. A gap of .19 to .34 exists between the easiest test and the next test in all of these configurations compared to virtually equal correlations for these two tests for all other configurations. For these five configurations, the loss of simulees (1150, 1412, 1336, 1442, 1079, respectively) is noticeably larger than nearly all other conditions for all other configurations. It appears that a loss of over 1000 examinees will noticeably lower the correlations between estimated thetas and true thetas on both dimensions.

Examination of the situations in which examinees are lost due to zero or perfect scores reveals other interesting results. From Table 12, the Group M1 configurations show a loss of 17 to 1442 simulees (.3-24.0%) with an average loss of 7.1%, 9.9%, 9.6%, 11.1%, and 7.4% for Configurations 1, and 7-10 respectively. Group M2 (Configuration 5) shows a loss of 7 to 694 simulees (.1-11.6%) with an average loss of 3.4% for the six conditions of Configuration 5. Group U shows a loss of 5 to 687 (.1-11.4%) with an average loss of 2.5%, 3.6%, and 3.3% for Configurations 2, 4, and 6, respectively. This is comparable to the unidimensional criterion which shows a loss of 8 to 569 (.1-9.5%) with an average loss of 3.0%.

Configuration 3 has an average loss of 4.2% over the six conditions but breaking it up into Test 1 (4.9%) versus Test 2 (3.4%) again allows a comparison of Test 1 with the other unidimensional tests and Test 2 with the other multidimensional tests. The Group U Test 1 average losses per test are all under 6.1% while the Group M1 are all over 11.9%. Configuration 3 Test 1 clearly falls in with the other unidimensional configurations, as would be expected. Group U Test 2 average losses are all under 1.1%, while the Group M1 are all over 2.2%. Configuration 3 Test 2 falls in with the multidimensional configurations as would be expected. Note that Configuration 3 is the only configuration with a large gap between N* for Test 1 $\bar{b}_2 - \bar{b}_1 = 0$ and Test 2 $\bar{b}_2 - \bar{b}_1 = 0$. The multidimensional test lost 356 more simulees than the unidimensional test.

Configuration 5 is again in a group by itself with Test 1 (6.3%) average loss being greater than all Group U configurations and less than all Group M1 configurations, and Test 2 falling within the Group U percentages. The average loss per all six conditions of Configuration 5 is 3.4%, which again falls within the Group U percentages. Although not as dramatic, Yen's (1984b) results using N = 1000 also show the drop in correlations between estimated theta and both true thetas, and the increase in examinees lost due to zero or perfect scores in the easiest test.

Clearly, in all cases except Configuration 5, the multidimensional tests lost more simulees to zero or perfect scores than the undimensional tests. This could have serious implications for item and ability parameter estimation.

How well both traits are estimated, then, appears to depend on how strong the correlation between true ability on the two traits is, on whether the two tests are unidimensional or multidimensional according to the "loading" criterion, and on how many examinees are lost due to zero or perfect scores. The better the correlation between true ability on the two traits, the better the estimation of both traits. (Of course, LOGIST is meant to estimate only one trait.) If one or both tests are considered to be multidimensional according to the "loading" criterion, then both traits are estimated fairly well; however, if both tests are unidimensional, then one trait is well estimated and one is estimated poorly. If over 1000 examinees are lost, the traits are poorly estimated, despite N (approximately 3000) being larger than the criterion set by previous researchers. These results support several studies where the same conclusions were reached (Christoffersson, 1975; Drasgow & Parsons, 1983; McKinley, 1983; Reckase, 1977, 1979; Yen, 1984d).

Reckase (1977, 1979) found that when there is a dominant first factor present in multidimensional data, then the three-parameter model estimates that single factor. The Group U data sets here clearly have a dominant first factor. Trait 1 is well estimated and Trait 2 is not well estimated. This is also true for Configuration 3 Test 1. These data set results support the Reckase findings. Just as clearly, the other configurations have nearly equal estimation of both traits and these estimates get better for both traits as the correlation between θ_1 and θ_2

increases. This does not support the Reckase findings. However, Reckase's (1977, 1979, 1981b) conclusion was that although unstable item parameter estimates may result, good ability estimates can be obtained despite the data being multidimensional. This conclusion is supported here. The descrepancies found here between the Reckase findings and those indicated by the data presented in this paper may very well be due to the different sample sizes (Reckase N = 1000), the different generating models (Reckase used a linear factor analysis model), and sampling error.

Drasgow and Parsons (1983) found that when one trait is sufficiently prepotent (dominant), then a unidimensional model provides a good description of multidimensional data sets. The results shown here for the Group U configurations support this conclusion. However, the conclusions from this data go beyond that, indicating that even with two-dimensional data, the trait estimates are good enough to conclude that a unidimensional model can describe multidimensional (two-dimensional) data well at least when the correlation between θ_1 and θ_2 is above .5.

Yen's (1984b) mathematical predictions support the hypothesis that multidimensional data analyzed by the unidimensional three-parameter model result in a unidimensional trait that is a combination of the underlying traits. If the test involves traits that influence all or most of the items the prediction is that the underlying true traits have approximately equal influence in determining estimated θ. Her simulated results confirm that prediction, as do the correlations of true and estimated thetas for Group M1 and M2 configurations here. If the test involves independent traits, one of which influences only a few items, that trait is ignored in the definition of the unidimensional three-parameter trait. Group U correlations support this second prediction as do the Reckase (1979) results.

Summary and conclusions. It is accepted knowledge that many existing standardized tests, such as most achievement tests and many aptitude tests, do not satisfy the undimensionality assumption of the three-parameter logistic model (Bejar, 1983; Bock; 1979; Hambleton & Cook, 1977; Hutten, 1980; Kingston & Dorans, 1982; Reckase, 1977, 1979). Therefore, the question to be answered is not whether the assumption is satisfied but whether a specific use of the model is robust to violations of the assumption (Hambleton & Cook, 1977; Hambleton, Swaminathan, Cook, Eignor, & Gifford, 1978; Reckase, 1981). Hambleton et al. (1978) and Yen (1984a, 1984b) presented evidence that the models are robust to some departures. The results of this research present more.

Factor analyses were used in order to verify whether the data were truly multidimensional or not. The factor analyses supported a division of the simulated multidimensional data sets into groups according to how the tests "load" (discriminate) on the two dimensions. The tests either both "load" heavily on both dimensions (both tests are multidimensional), one test "loads" heavily on one dimension and the other test "loads" heavily on the same dimension (both tests are unidimensional), one test is unidimensional and one multidimensional, or one test "loads" heavily on one dimension and the other test loads heavily on the other dimension.

Although the strength of the correlation between the two generating traits seemed to have little effect on the quality of the parameter estimation, there is evidence that the unidimensionality or multidimensionality of the tests, as determined by both factor analyses and the discrimination loadings on the two dimensions, does have an effect on item parameter estimation.

When both tests are unidimensional, both loading most heavily on the same dimension, then a_1 and b_1 are well estimated, and a_2 and b_2 are poorly estimated. If both tests to be equated are multidimensional, then b_1 is estimated fairly well, b_2 is poorly estimated, and a_1 and a_2 are mostly poorly estimated. If both tests are multidimensional with Test 1 loading heavily on one dimension and Test 2 loading heavily on the other dimension, then b_1 and b_2, a_1 and a_2 are all poorly estimated. If Test 1 is unidimensional and Test 2 is multidimensional, then for Test 1 a_1 and b_1 are well estimated and a_2 and b_2 are poorly estimated, while for Test 2 b_1 is fairly well estimated, and b_2, a_1, and a_2 are poorly estimated. The estimation of the c parameter was mediocre to poor for all conditions of all configurations.

The results of this research indicate that the poorest item parameter estimates occur for the situation in which one test is unidimensional and one is multidimensional, such as a situation in which Trait 2 is measured by only a few items on one test and by most or all of the items on the other test. This situation appears to be worse than if both tests are unidimensional or both are multidimensional.

In conclusion, these results indicate that use of the three-parameter logistic model is as good, in most instances, for parameter estimation of multidimensional data as it is for unidimensional data for the types of conditions studied in this research. Caution should be exercised, however, when one test is unidimensional and one is multidimensional, such as occurs when a higher level has a few items measuring a trait that a lower level does not measure.

REFERENCES

Bejar, I. I. (1983). Introduction to item response models and their assumptions. In R. K. Hambleton, (Ed.), Applications of item response theory. Vancouver, BC: Educational Research Institite of British Columbia.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. Novick, Statistical theories of mental test scores. (pp. 395-549) Reading, Mass.: Addison-Wesley.

Bock, R. D. (1979). A feasibility study of the one-, two-, and three-parameter item response models for the analysis and reporting of the California Assessment Program. Chicago: International Educational Services.

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

Comprehensive Tests of Basic Skills, Forms U & V (1981) Monterey, CA: CTB/McGraw-Hill.

Cook, L. L. & Eignor, D. R. (1981, April). Score equating and item response theory: Some practical considerations. Paper presented at annual meeting of the National Council on Measurement in Education, Los Angeles.

Doody-Bogan, E., & 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.

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

Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 14, 75-96.

Hambleton, R. K., & Martois, J. S. (1983). Evaluation of a test score prediction system based upon item response model principles and procedures. In R. K. Hambleton (Ed.), Applications of item response theory. Vancouver, BC: Educational Research Institute of British Columbia.

Hambleton, R.K., Swaminathan, H., Cook, L.L., Eignor, D.R., & Gifford, J.A. (1978). Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 48, 467-510.

Hattie, J. (1982). Decision criteria for determining unidimensionality: An empirical study. Armidale, N.S.W., Australia: Center for Behavioral Studies.

Hutten, I. R. (1980, April). Some empirical evidence for latent trait model selection. Paper presented at the annual meeting of the American Educational Research Association, Boston.

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

Kingston, N. M. & Dorans, N.J. (1982). Equating of GRE aptitude tests using item response theory: Effects of multidimensionality. Paper presented at the annual meeting of the American Psychological Association.

Lord, F.M. (1968) Analysis of Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 28, 989-1020.

Lord, F. M. (1975). Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters (Research Bulletin 75-33). Princeton, NJ: Educational Testing Service.

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

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. Journal of Educational Measurement, 17, 179-193.

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.

McKinley,R.L. & Reckase,M.D. (1980, December). A comparison of the ANCILLES and LOGIST parameter estimation procedures for the three-parameter logistic model using goodness of fit as a criterion. (Research Report 80-2). Columbia: University of Missouri, Department of Educational Psychology.

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

McKinley, R. L., & Reckase, M. D. (1983a, August). An extension of the two-parameter logistic model to the multidimensional latent space Research Report ONR 83-2). Iowa City, IA: The American College Testing Program.

McKinley, R. L., & Reckase, M. D. (1983b, August). An application of a multidimensional extension of the two-parmeter logistic latent triat model Research Report ONR 83-3). Iowa City, IA: American College Testing Program.

McKinley, R. L., & Reckase, M. D. (1984, June). Correlated abilities in multidimensional item response theory. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Reckase, M.D. (1977). 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.

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. (1981a, April). The validity of latent trait models through the analysis of fit and invariance. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

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

Reckase, M. D., & McKinley, R. L. (1982). Some latent trait theory in a multidimensional latent space. Paper presented at the 1982 Invitational IRT/CAT Congerence, Wyzata, Minnesota.

Ree, M. J. & Jensen, H. E. (1983). Effects of sample size on linear equating of item characteristic curve parameters. In D. J. Weiss (Ed.), New horizons in testing. New York: Academic Press.

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

Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), Applications of item response theory. VAncouver, BC: Educational Research Institute of British Columbia.

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

Yen, W. M. (1980). The extent, causes and importance of context effects on item parameters for two latent trait models. Journal of Educational Statistics, 17, 297-311.

Yen, W.M. (1983). Tau equivalence and equipercentile equating. Psychometrika, 48, 353-369.

Yen, W. M. (1984a). Effects of local item dependence on the fit and equating perfor ance of the three-parameter logistic model. Applied Psychological Measurement, 8, 125-145.

Yen, W. M. (1964b, June). Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory. Paper presented at the Psychometric Society, Santa Barbara, CA.

Yen, W. M. (1984c). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. Journal of Educational Statistics, 21, 93-111.

Yen, W. M. (1984d, April). The choice of scale for educational measurement: An IRT perspective. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.