Psicothema, 1999. Vol. 11, nº 1, pp. 175-182
ISSN 0214 - 9915 CODEN PSOTEG

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

ROBUSTNESS OF ITEM RESPONSE LOGISTIC MODELS TO VIOLATIONS OF THE UNIDIMENSIONALITY ASSUMPTION

Marcelino Cuesta y José Muniz
Universidad de Oviedo

The effects of violating the unidimensionality assumption when applying item response logistic models were studied. Using a multidimensional two parameter logistic model, two different tests that are common in practice were simulated: a) a test composed of two equally relevant dimensions, and b) a test with a dominant dimension and a secondary one. Each test was composed of 40 items, 25 corresponding to the first dimension, 15 to the second. Two sample sizes (N= 300 and N= 1000) and five levels of correlation between the dimensions (0.05, 0.30, 0.60, 0.90, 0.95) were used to generate the data. The unidimensional two-parameter logistic model was used to estimate the item parameters and the ability of the examinees. The results indicate that the unidimensional estimates are consistently robust. Estimates of the item difficulty parameter are less affected by the violation of the unidimensionality assumption than the other item parameter estimates. The item discrimination parameter and the ability estimates are influenced by the size of the correlation between the dimensions and by the type of multidimensionality displayed by the data.

Robustez de los modelos logísticos de respuesta a los ítems a las violaciones del supuesto de unidimensionalidad. Se estudiaron los efectos de violar la asunción de unidimensionalidad cuando se aplican modelos logísticos de teoría de respuesta a los items. Usando un modelo logístico multidimensional de dos parámetros, se simularon dos tipos distintos de tests de uso común: a) un test compuesto de dos dimensiones igualmente relevantes y b) un test con una dimensión dominante y otra secundaria. Cada test se componía de 40 items, 25 de los cuales correspondían a la primera dimensión y 15 a la segunda. En la generación de los datos se emplearon dos tamaños de muestra (N=300 y N=1000) y cinco niveles de correlación entre dimensiones (0.05, 0.3, 0.6, 0.9 y 0.95). El modelo logistico unidimensional de dos parámetros fue usado para estimar los parámetros de los items y la habilidad de los sujetos. Los resultados indican que las estmaciones unidimensionales son consistentemente robustas. Las estimaciones del parámetro de dificultad se ven menos afectadas por la violación del supuesto de unidimensionalidad que el resto de los parámetros estimados. El parámetro de disciminación de los items y la habilidad estimada se ven influenciados por el tamaño de la correlación entre dimensiones y por el tipo de multidimensionalidad mostrado por los datos.

Correspondencia: Marcelino Cuesta
Facultad de Psicología
Universidad de Oviedo
33003 Oviedo (Spain)
E-mail: mcuesta@sci.cpd.uniovi.es

The most widely used item response theory models (the one-, two- and three-parameter logistic models) require that the data be unidimensional. With the growing utili-

zation of these models to evaluate psychological and educational variables, some questions about their proper use have arisen. One of the most important issues to arise has been that of the nonfulfillment of the unidimensionality assumption by most of the real data analyzed with these models. Different authors have pointed out the difficulty of finding psychological and educational variables which strictly meet the condition of unidimensionality (Harrison, 1986; Hulin, Drasgow & Parsons, 1983; Reckase, 1979, 1985, 1989; Reckase & McKinley, 1982).

In addition to the nature of the construct being measured by a test, other collateral aspects can influence dimensionality. Birenbaum and Tatsuoka (1982) pointed out the effect of instruction on test dimensionality, and Traub (1983) states some questions which can affect the dimensionality of the test, such as instructions, the speed conditions, or the tendency of examinees to guess. Rosenbaum (1988) also poses the possibility of the presence of item bundles which share some information in common and could violate the assumption of unidimensionality. This preoccupation about the dimensionality of the real test data gave rise to a line of investigation centered on the robustness of item response theory (IRT) models to the violation of the unidimensionality assumption. Reckase (1979) found that when the dimensions of a test were equally important, the ability estimates obtained using a unidimensional model represented the average of the dimensions. Whereas, when the test was composed of a dominant dimension and a secondary one (Stout, 1987), unidimensional estimates of ability tend to capture the first factor. Within this framework, Drasgow and Parsons (1983) suggested avoiding the use of unidimensional models when the correlation between dimensions is below 0.40. Harrison (1986) and Cuesta and Muñiz (1994, 1995) reported similar results. Using two-dimensional data, Ansley and Forsyth (1985) found that the unidimensional estimates of discrimination and ability parameters tend to approximate the average of both dimensions, whereas the unidimensional estimates of the difficulty parameter seem to overestimate the parameters of the first dimension. Way, Ansley, and Forsyth (1988) did not find relevant differences when using compensatory and non-compensatory models. Results converging in the same direction are reported by other authors too (Ackerman, 1989; Doody-Bogan & Yen, 1983; McKinley & Reckase, 1983; Reckase, 1985; Yen, 1984).

In general terms, most of these studies show that IRT logistic models appear to be robust to moderate violations of the unidimensionality assumption. However, many situations common to everyday testing practice, in which the assumption of data unidimensionality is not likely to be strictly fulfilled, remain to be investigated. The central aim of this paper is to investigate the behavior of item and ability parameter estimates obtained with a unidimensional two-parameter logistic model when the data are bidimensional.

Two common situations in testing practice were investigated: *Case 1*, a test with two dimensions which are equally dominant, and *Case 2*, a test with one dominant dimension and one secondary one. These are two situations many practitioners face every day when evaluating psychological and educational traits. The correlations between the two dimensions were also taken into account.

Method

Data Simulation

The model used to simulate the data (McKinley & Reckase, 1983) is a multidimensional extension of the two-parameter logis-

tic model. According to this compensatory model, the probability of a correct response to an item is:

$$P(x_{ij}=1/a_i,d_i,\theta_j)= [e^{(a'_i\theta_j+d_i)}]/[1+e^{(a'_i\theta_j+d_i)}] \quad (1)$$

where:

$P(x_{ij}=1/a_i,d_i,\theta_j)$ is the probability of a correct response to item i by examinee j,

a_i is a discrimination parameter vector,

d_i is a parameter related to the difficulty of the item, and

θ_j is an ability parameter vector.

The exponent of the previous expression can be rewritten as

$$\sum_{k=1}^{n} a_{ik}\,(\theta_{jk} - b_{ik}), \quad (2)$$

where:

n the number of dimensions,

a_{ik} an element of a_i,

θ_{jk} an element of θ_j, and

$$d_i = -\sum_{k=1}^{n}(a_{ik}b_{ik})$$

Based on the McKinley and Reckase (1983) model, Reckase (1985) proposed a new approach to the concept of difficulty, represented by d_i in the former model. Reckase introduces the multidimensional item difficulty (MID), which corresponds to the point of the item response surface (IRS) where the item has the highest discriminatory power, that is to say where the item information is a maximum. In the unidimensional case, this value is given by the point of greatest slope of the item characteristic curve. However, when more than one dimension is involved, to define the difficulty parameter, the slope of a given point depends on the direction under consideration. That is why Reckase uses the distance from the origin of the latent space to the point of maximum discrimination, as well as the direction of this point with respect to the axes representing the dimensions under consideration. The distance from the origin is calculated according to the following expression:

$$D_i = \frac{-d_i}{(\sum_{k=1}^{n} a_{ik}^2)^{1/2}} \quad (3)$$

and the direction:

$$\cos \alpha_{ik} = \frac{a_{ik}}{(\sum_{k=1}^{n} a_{ik}^2)^{1/2}} \quad (4)$$

In accordance with this redefinition of the difficulty, for two items to be comparable it is necessary that they measure the same combination of abilities, that is to say, that they have the same direction.

Using MID as a starting point, Reckase (1986) proposed a related multidimensional discrimination index (MDISC). The definition put forward by the MDISC is presented as a function of the slope of the IRS at the point of greatest slope, in the direction indicated by the MID. The value of this parameter is:

$$MDISC = (\sum_{k=1}^{n} a_{ik}^2)^{1/2} = \frac{-d_i}{MID_i} \quad (5)$$

The generation of data according to this multidimensional model was performed with the M2GEN2 program developed by Ackerman (1989). The program allows the generation of two-dimensional data with

different levels of correlation. As input, this generator requires a discrimination parameter vector for each of the dimensions, and a vector of item difficulties. As output, the program offers the examinees ability for each of the two dimensions, and the matrix (examinees *x* items) of ones and zeros from which the values were estimated.

Data Sets

Five levels of correlation between the generated dimensions were used: 0.05, 0.30, 0.60, 0.90, 0.95. Two sample sizes were used: N=300 and N=1000. The two generated dimensions, θ_1 and θ_2, were scaled with a mean of zero and variance of one, N(0,1).

To simulate the ability parameters (*Case 1*), the Reckase (1985, 1986) data were used: 25 highly discriminating items on one of the dimensions, and 15 on the other (Table 1). The same values were used for *Case 2*, but the highest discrimination indices always appeared on the first dimension.

Table 1
Discrimination indices of the items

Item	a_{i1}	a_{i2}	Item	a_{i1}	a_{i2}
1	1.81	0.86	21	1.41	0.04
2	1.22	0.07	22	1.54	1.79
3	1.57	0.36	23	0.54	0.23
4	0.71	0.53	24	1.53	0.48
5	0.86	0.19	25	0.72	0.55
6	1.72	0.18	26	0.51	0.65
7	1.86	0.29	27	1.66	1.72
8	1.33	0.34	28	0.69	0.19
9	1.19	1.57	29	0.88	1.12
10	2.00	0.00	30	0.68	1.21
11	0.87	0.00	31	0.24	1.14
12	2.00	0.98	32	0.51	1.21
13	1.00	0.89	33	0.76	0.59
14	1.22	0.14	34	0.01	1.94
15	1.27	0.47	35	0.39	1.77
16	1.35	1.15	36	0.76	0.99
17	1.06	0.45	37	0.49	1.10
18	1.92	0.00	38	0.29	1.10
19	0.96	0.22	39	0.48	1.00
20	1.20	0.12	40	0.42	0.75

Data Analysis

Coefficient alpha (Cronbach, 1951), factor analysis, and other descriptive statistics were performed using the SPSS/PC statistical package. Logistic item response models parameter estimates were obtained via BILOG (Mislevy & Bock, 1984). The root mean squared differences (RMSD) and Pearson's correlations were used to compare the multidimensional parameters with the correspondent unidimensional estimates.

Results

Case 1

Table 2 shows the descriptive statistics for the data for case 1. The correlations between the ability simulated data [$r(\theta_1,\theta_2)$] are very close to the levels of correlation intended. The coefficient alpha appears to be high under all conditions, ranging from 0.90 to 0.95. The first three eigenvalues obtained from a principal component analysis are reported as an approximation to the dimensionality of each set of test data. The explained variance increases with the correlation between the dimensions. The mean and standard deviation are also reported. The sample sizes used (N= 300, and N= 1,000) did not seem to play an important role in the accuracy of the estimates, that is why in the ta-

Table 2
Descriptive statistics of the test data: Case 1

	Correlation between the dimensions of the test				
Statistic	0.05	0.30	0.60	0.90	0.95
$r(\theta_1,\theta_2)$	.04	.31	.69	.98	.99
Mean	17.87	18.00	18.08	18.05	18.04
SD	8.96	9.54	10.10	10.80	10.98
α	.91	.92	.93	.94	.95
λ_1	9.11	10.22	11.40	12.87	13.29
λ_2	2.53	2.11	1.54	1.13	1.13
λ_3	1.10	1.06	1.07	1.07	1.05

bles only the results corresponding to N= 1,000 are reported.

Table 3 shows the accuracy of unidimensional model parameter estimates obtained from two-dimensional data. As the correlation between dimensions increases, the precision of the estimates improves. The discrimination for an item parameter on the second dimension (a_2) is always closer to the unidimensional estimate (a') than the discrimination parameter for the item on the first dimension (a_1). However, the mean of both item discrimination parameters (a_m) is closest to the unidimensional estimate. The greatest distance appears with respect to the multidimensional discrimination index (*md*).

Table 3
Root mean squared differences (RS) and correlations (r) between item parameters: Case 1

$\rho(\theta_1,\theta_2)$	RS(a_1-a')	RS(a_2-a')	RS(a_m-a')	RS(md-a')	RS(d-b')	RS(D-b')
0.05	.54	.51	.23	.78	1.61	.19
0.30	.52	.49	.17	.72	1.51	.13
0.60	.53	.45	.13	.68	1.45	.14
0.90	.50	.45	.06	.58	1.40	.14
0.95	.49	.47	.06	.57	1.40	.15
$r(\theta_1,\theta_2)$	$r(a_1,a')$	$r(a_2,a')$	$r(a_m,a')$	r(md,a')	r(d,b')	r(D,b')
0.05	.69	.39	.93	.83	-.96	.97
0.30	.63	.48	.96	.87	-.96	.98
0.60	.50	.62	.97	.88	-.95	.98
0.90	.47	.66	.98	.91	-.95	.99
0.95	.50	.63	.98	.92	-.95	.98

The parameters a_1 and a_2 follow inverse patterns in their correlation with the unidimensional estimates. When the correlations between the dimensions are lower, the unidimensional estimates are more correlated with the item discrimination parameter of the first dimension. As the correlation between both dimensions increases, the correlation of the unidimensional estimates with the second dimension increases, decreasing with the first.

Very high correlations were found between the unidimensional difficulty parameter estimates and parameters *d* and *D*.

The accuracy of the unidimensional ability estimates increased with the increasing of correlation between dimensions (Table 4). These estimates are very close to the mean of the ability parameters of both dimensions (θ_m), with correlations ranging from 0.93, when dimensions are uncorrelated, to 0.97, when the dimensions correlate 0.98.

Table 4
Root mean squared differences (RS) and correlations (r) between ability parameters: Case 1

$\rho(\theta_1,\theta_2)$	RS(θ_1-θ')	RS(θ_2-θ')	RS(θ_m-θ')	$r(\theta_1,\theta')$	$r(\theta_2,\theta')$	$r(\theta_m,\theta')$
0.05	.64	.95	.42	.79	.56	.93
0.30	.56	.78	.36	.84	.70	.95
0.60	.47	.55	.35	.90	.86	.96
0.90	.31	.31	.30	.96	.96	.96
0.95	.29	.29	.29	.96	.96	.96

Case 2

In this second case, as previously pointed out, a test with a principal dimension and a secondary one was simulated. The descriptive statistics of the data used appear in Table 5.

Table 5
Descriptive statistics of the test data: Case 2

	Correlation between the dimensions of the test				
Statistic	0.05	0.30	0.60	0.90	0.95
$r(\theta_1,\theta_2)$	.04	.31	.69	.98	.99
Mean	17.83	17.97	18.04	18.01	18.00
SD	9.44	9.81	10.13	10.72	10.93
α	.92	.93	.93	.94	.95
λ_1	10.07	10.80	11.47	12.71	13.17
λ_2	1.42	1.30	1.20	1.12	1.12
λ_3	1.15	1.10	1.09	1.07	1.05

As in case 1, the values of coefficient alpha are very high for all data bases. The

size of the first eigenvalue, and the similarity of the second and the third, indicates in all cases factorial unidimensionality. It seems, therefore, that from a factorial point of view this test is clearly unidimensional.

The estimation of the discrimination parameter (see Table 6) is closer to the mean of the parameters of the dimensions than to any of the other indicators considered. The multidimensional discrimination index also has high correlations with the unidimensional estimates.

Table 6
Root mean squared differences (RS) and correlations (r) between item parameters: Case 2

$\rho(\theta_1,\theta_2)$	RS(a_1-a')	RS(a_2-a')	RS(a_m-a')	RS(md-a')	RS(d-b')	RS(D-b')
0.05	.60	.42	.21	.73	1.56	.10
0.30	.56	.42	.16	.68	1.50	.09
0.60	.55	.42	.12	.66	1.46	.12
0.90	.49	.46	.05	.59	1.41	.14
0.95	.47	.48	.05	.57	1.40	.14
$r(\theta_1,\theta_2)$	$r(a_1,a')$	$r(a_2,a')$	$r(a_m,a')$	r(md,a')	r(d,b')	r(D,b')
0.05	.87	.52	.92	.94	-.94	.99
0.30	.84	.61	.95	.94	-.95	.99
0.60	.78	.61	.97	.92	-.95	.99
0.90	.74	.75	.98	.91	-.95	.99
0.95	.76	.74	.99	.92	-.96	.99

The correlations between the *b* estimates and the parameters *d* and *D* presented in Table 7 were very high (|0.92 to 0.99|). As regards the estimation of the ability (Table 7) of the subjects, the predominance of the first dimension over the second is clear. As the correlation between the two dimensions increases, the relation between the unidimensional estimation and the second dimension also increases. Correlations between the unidimensional estimates and the ability average of both dimensions are very high, with values ranging from 0.88 to 0.97.

Table 7
Root mean squared differences (RS) and correlations (r) between ability parameters: Case 2

$\rho(\theta_1,\theta_2)$	RS(θ_1-θ')	RS(θ_2-θ')	RS(θ_m-θ')	$r(\theta_1,\theta')$	$r(\theta_2,\theta')$	$r(\theta_m,\theta')$
0.05	.48	1.59	.51	.89	.40	.88
0.30	.44	.93	.45	.91	.59	.91
0.60	.40	.65	.38	.94	.81	.94
0.90	.30	.33	.31	.96	.95	.96
0.95	.29	.30	.29	.96	.96	.96

Conclusions

The main goal of this research was to investigate the degree to which the violation of the unidimensionality assumption affects the applicability of the most popular item response logistic models. Of the three unidimensional model parameters which have been considered (*a*, *b*, and q), the difficulty (*b*) seems to be the least affected by the violation of the unidimensionality assumption. This result is in accordance with that found in similar works by Ackerman (1989, 1991) and Oshima and Miller (1990). The parameter *d*, as well as the distance to the point of maximum discrimination, are seen to be highly related to the unidimensional item difficulty estimates, exhibiting no important differences between the results obtained in case 1 and case 2. The unidimensional estimates of the discrimination parameter *a* seem to capture the average values of the parameters assigned to each of the two dimensions when there exists a certain correlation between the dimensions ($r \geq 0.3$). The correlations between the unidimensional estimates and MDISC are also high; especially when the dimensions are strongly correlated. The correlations between MDISC and the unidimensional estimates (case 2) always have higher values than those found in case 1. Some differences are observed between the estimates of the item discrimination parameter. While in the first

case, a_1 initially captures the attention of the unidimensional estimates, that attention gradually changing to a_2; in the second case, a relatively high correlation with a_1 is always present, converging also a_2 to this correlation when the relation between θ_1 and θ_2 increases. In the tests with uncorrelated dimensions, it was found that the relation between a_1 and a_2 with the unidimensional estimates is very close. The sample sizes used (N=300 and N=1.000) do not seem to affect the accuracy of estimates. The robustness of the model parameter estimates is especially strong when the correlation between the two dimensions of the simulated test is above 0.30; increasing the precision with increasing correlation between the dimensions. The unidimensional estimates of bidimensional tests tend to capture the average of the parameters of the test dimensions.

Confirming most of the previous research (Ackerman, 1989; Ansley & Forsyth, 1985; Drasgow & Parsons, 1983; Harrison, 1986; Way, Ansley & Forsyth, 1988; Yen, 1984), the general conclusion of this study is that the unidimensional estimates of item parameters (difficulty and discrimination) and examinee ability were consistently robust to moderate violations of test unidimensionality. At an applied level, these results seem to indicate that when the test has a dominant dimension, even when the dimensions measured by the test are uncorrelated, the violation of the assumption of unidimensionality does not produce serious errors in model parameter estimation, especially with respect to ability estimation.

Referencias

Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement, 13*(2), 113-127.

Ackerman, T. A. (1991). The use of unidimensional parameter estimates of multidimen sional items in adaptative testing. *Applied Psychological Measurement, 15*(1), 13-24.

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

Birenbaum, M., & Tatsuoka, K. K. (1982). On the dimensionality of achievement test data. *Journal of Educational Measurement, 19*(4), 259-266.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cuesta, M., & Muñiz, J. (1994). Utilización de modelos unidimensionales de teoría de respuesta a los items con datos multifactoriales. *Psicothema, 6*(2), 283-296.

Cuesta, M., & Muñiz, J. (1995). Efectos de la multidimensionalidad en la estimación de parámetros desde modelos unidimensionales de teoría de respuesta a los items. *Psicológica, 16*, 65-86.

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

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory model to multidimensional data. *Applied Psychological Measurement, 7*(2), 189-199.

Harrison, D. A. (1986). Robustness of IRT paremeter estimation to violations of unidimensionality assumption. *Journal of Educational Statistics, 11*(2), 91-115.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Applications to psychological measurement.* Homewood, IL: Dow Jones-Irvin.

McKinley, R. L., & Reckase, M. D. (1983). *An extension of the two parameter logistic model to the multidimensional latent space* (Research Report No. 83-2). Iowa City, IA: The American College Testing Program.

Mislevy, R. J., & Bock, R. D. (1984). *BILOG: Maximum likelihood item analysis and test scoring with logistic models*. Mooresville, IN: Scientific Software.

Oshima, T. C., & Miller, M. D. (1990). Multidimensionality and IRT based item invariance indexes: The effect of between-group variation in trait correlation. *Journal of Educational Measurement, 27*, 273-283.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics, 4*(3), 207-230.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*(4), 401-412.

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

Reckase, M. D. (1989, August). *Controlling the psychometric snake: or, how I learned to love multidimensionality*. Invited address at the meeting of the American Psychological Association, New Orleans.

Reckase, M. D., & McKinley, R. L. (1982). *The feasibility of a multidimensional latent trait model.* Paper presented at the meeting of the American Psychological Association, Washington.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*(4), 589-617.

Rosenbaum, P. R. (1988). Items bundles. *Psychometrika, 53*, 349-359.

Traub, R. E. (1983). A priori considerations in choosing an item response model. In R. K. Hambleton (Ed.), *Applications of item response theory*. (pp. 57-70). British Columbia: Educational Research Institute of British Columbia.

Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement, 12*(3), 239-259.

Yen, W. M. (1984). Effects of local dependence on the fit and equating performance of the three parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145.

Aceptado el 15 de junio de 1998