Psicológica (2014), 35, 149-169.

Detection of Q-matrix misspecification using two criteria for validation of cognitive structures under the Least Squares Distance Model

Sonia J. Romero[1], Xavier G. Ordoñez[2], Vicente Ponsoda[3], and Javier Revuelta[3]

[1]*Universidad a Distancia de Madrid, Spain*

[2]*Universidad Complutense de Madrid, Spain*

[3]*Universidad Autónoma de Madrid, Spain*

Cognitive Diagnostic Models (CDMs) aim to provide information about the degree to which individuals have mastered specific attributes that underlie the success of these individuals on test items. The **Q**-matrix is a key element in the application of CDMs, because contains links item-attributes representing the cognitive structure proposed for solve the test. Using a simulation study we investigated the performance of two model-fit statistics (MAD and LSD) to detect misspecifications in the **Q**-matrix within the least squares distance modeling framework. The manipulated test design factors included the number of respondents (300, 500, 1000), attributes (1, 2, 3, 4), and type of model (conjunctive vs disjunctive). We investigated MAD and LSD behavior under correct **Q**-matrix specification, with **Q**-misspecifications and in a real data application. The results shows that the two model-fit indexes were sensitive to **Q**-misspecifications, consequently, cut points were proposed to use in applied context.

Cognitive Diagnostic Models (CDMs) aim to provide information about the degree to which individuals have mastered specific attributes (e.g., cognitive operations and processes or skills) that underlie the success of these individuals on test items. Interest in CDMs has been growing

[1] Correspondence concerning this article should be addressed to Sonia J. Romero, Department of Psychology, Universidad a Distancia de Madrid, Carretera de la Coruña Km 38,5, vía de servicio n° 15, Collado Villalba, 28400 Madrid, Spain. Email: soniajaneth.romero@udima.es

rapidly over the past years and many different models have been developed (e.g., de la Torre, 2009; de la Torre & Douglas, 2004; DiBello, Stout, & Roussos, 1995; Dimitrov, 2007; Dimitrov & Atanasov, 2011; Henson & Douglas, 2005; Junker & Sijtsma, 2001; Templin & Henson, 2006). Some of these models are based on log-linear models (Henson, Templin & Willse, 2009), other on deterministic inputs, noisy and gate models (de la Torre, 2009; de la Torre & Douglas, 2004), others are more general (von Davier, 2010) and some are based on Item Response Theory (IRT; Dimitrov & Atanasov, 2011; Dimitrov, 2007; Embretson, 1984, 1993; Fisher, 1995; Tatsuoka, 1985, 1995); this is the case of the Least Squares Distance Model (LSDM) proposed by Dimitrov (2007).

As any other CDM, LSDM requires the user to develop a **Q**-matrix specifying which attributes are required by each item; and then uses the item parameters estimated with the IRT model and the **Q**-matrix to estimate the probability of correct performance on each attribute, across fixed ability levels. The validity of the results depends on the correct specification of the **Q**-matrix (Corter, 1995; de la Torre, 2008; Dimitrov & Raykov, 2003; Medina-Diaz, 1993; Romero, 2010). Incorrect specification of the **Q**-matrix leads to misclassifications of the examinees in the latent classes (Rupp & Templin, 2008) and, consequently, to erroneous diagnosis in the attribute mastery.

The correctness of the **Q**-matrix is a very important issue to evaluate in practice. One of the methods proposed to address this is LSDM. This method allows the validation and analysis of cognitive attributes required for correct answers of binary items across fixed ability levels. The principal advantage of LSDM is that not require information about the score of examinees if the IRT parameters are available. In addition, the model permits Q-validation screening previous to test administration.

Despite its importance, there has been little research in the context of CDMs about how sensitive are different model-data fit indexes to different data conditions and model misspecification. For that reason, we conducted a simulation to evaluate the behavior and distribution of two LSDM statistical fit indices with a correct **Q**-matrix specification. Based on the results of this study we propose cutoff points that can be used in applied context. Then, we report on a second simulation study evaluating the effects of model misspecification on: a) item parameter recovery, b) attribute behavior and c) the assessment of model fit using the two fit indices. Finally, we present an illustration with real data. The specific objectives of the present study are:

1. To describe the statistical distribution of the two indices, under several conditions, using correct **Q**-matrix specification, and, based

on these distributions, propose cut points to evaluate possible **Q**-misspecifications.

2. To study the capacity of the two indices for the detection of **Q**-matrix misspecifications under several conditions, varying the sample size, number of attributes and type of model.
3. To illustrate the use of indexes and cut points in an empirical study with real data.

The principal questions that motivate the present research are: a) what are the empirical sampling distributions of two LSDM indices under different simulation conditions?, b) how do the fit indices perform under model misspecification, once appropriate cut-offs from the empirical sample distributions are used? c) how sensitive are the fit indices on a real data application?.

The outline of this paper is as follows: In the next section we provide a brief overview of the least squares distance framework. In the next section we describe the design of the simulation studies, the relevant outcome measures of interest and some analytical approaches. Finally, an illustration of the use of the proposed indices is presented for a mathematic test applied to a sample of 2897 students of fourth grade in several Spanish schools.

LSDM

LSDM is a model for validation of cognitive structures and analysis of binary items using their IRT parameters. This method use the parameters estimated with a IRT model and the **Q**-matrix for estimate the probability of mastering attributes (A_k), in pre-fixed ability levels (logit scale). Like most cognitive diagnosis models, LSDM assumes a conjunctive relation between attributes, in other words, the probability of correct answer to an item is the product of the likelihood of all attributes required by that item:

$$P_{ij} = \prod_{k=1}^{K} \left[P(A_k = 1 \mid \theta_i)\right]^{q_{jk}} \qquad (1)$$

where, P_{ij} is the probability of correct item response for a person at ability level θ_i,, $P(A_k = 1 \mid \theta_i)$is the probability of correct performance in attribute A_k for a examinee at the ability level θ_i and q_{jk} is the element of the **Q**-matrix for item *j* and attribute A_k. The model expressed in equation (1) assumes that the mastery of an attribute is statistically independent for an examinee in a fixed ability level. The LSDM estimate the attribute probabilities directly using least squares (Lawson & Hanson, 1974).

Specifically, if we take the natural logarithm on both sides of (1), this leads to a system of equations of the form: **L**=**QX**, where, **L** is a (known) vector with elements ln(Pij), and **X** is the (unknown) vector with elements X_k =lnP(A_k=1|θ_i). Solutions for vector **X** are sought to minimize the Euclidian norm of the vector ||**QX**-**L**||. Given the solutions for **X**, the probability of correct performance on attribute A_k for a person with ability θ_i is found directly as the exponent of **X**. This is done for multiple ability values to obtain the probability curve for each attribute across fixed ability levels on the logit scale. Dimitrov (2007) calls these curves Attribute Probability Curves (APCs).

More recently, Dimtrov & Atanasov (2011) proposed a disjunctive version of LSDM, in which the correct response on an item may occur when at least one of the attributes associated with the item is correctly applied, in this case the probability of correct item response is estimated by:

$$P_{ij} = 1 - \exp\left\{\sum_{k=1}^{k} q_{jk} \ln\left[1 - P(A_k = 1 \mid \theta_i)\right]\right\} \qquad (2)$$

Under the conjunctive LSDM (LSDM-C) to obtain the correct response of an item is necessary to master all attributes associated with the item, while in the disjunctive version (LSDM-D), the correct response is obtained if at least one of the attributes associated with the item is mastered.

Dimitrov (2007) proposed two validation indices for the cognitive structure expressed in the **Q**-matrix:

a) Least Squares Distance (LSD): LSD is an ability-level fit index, corresponding to the residual after the minimization of the norm ||**QX**-**L**||, therefore, there are one LSD value for each of the fixed ability levels. LSD indicates accuracy of the vector **X** estimation. The smaller the LSD at a given ability level, the better the attributes holds (jointly for all the items).

Some theoretical properties of the LSD are presented in Appendix 1.

b) Mean Absolute Difference (MAD): is an item-level fit index, corresponding to the mean of the absolute differences (across the ability levels), between the IRT probabilities (ICCs) and its LSDM recovery (product of the attribute probabilities according to the **Q**-matrix). A good recovery of the ICC indicate that the attribute relations specified in **Q** explain properly this item and, conversely, a bad recovery indicates an incorrect specification of **Q** for this particular item. Dimitrov (2007) suggested that MAD < 0.05 indicates an adequate ICC recovery, $0.05 <$ MAD < 0.10 indicates a tolerable ICC recovery, and MAD ≥ 0.10 indicates a poor ICC recovery. These values are somewhat arbitrary because little is known about the empirical sample distribution of MAD.

METHODOLOGY

Two simulation studies were conducted to gain information about the values of MAD and LSD. Study I is aimed to: (a) analyze the distribution of MAD and LSD for both models (LSDM-C and LSDM-D) under a **Q**-matrix correctly specified and (b) provide appropriate cut-offs from these empirical sample distributions. Study II analyze the performance (on several conditions) of both indices under **Q**-matrix misspecification using the cut-offs established in Study I.

Procedure for Study I

The first task is to generate data under LSDM. This is done in seven steps.

1. An arbitrary **Q**-matrix of 15 items and 4 attributes was defined (see Table 1).
2. Four couples of arbitrary parameters (*a, b*) were used to generate the probabilities of mastering each attribute (**X** vector) with the 2PL IRT model.
3. Taking **Q** and **X** from steps 1 and 2, LSDM was applied in order to find **L** on equation **L** = **QX**.
4. By having the probabilities **L**, the next step is to determine the 15 couples of 2PL parameters (*a,b*) that minimize the difference between **L** and the IRT probabilities in order to reduce to minimum the discrepancy between IRT model and the LSDM. The item parameters found by the minimization (referred here as the "true" IRT parameters) are presented in Table 2. As expected, MAD and LSD values obtained after the LSDM analyses (using these "true" parameters) are close to zero (see Tables 3 and 4).
5. The "true" IRT item parameters were used to simulate 500 data sets with the responses of 300, 500 or 1000 examinees to the 15 binary items. These data were then calibrated with the 2PL; (the mean of calibrated parameters on each condition are presented in Table 2).
6. Using the parameters from these calibrations (and the true **Q**-matrix), the LSDM was performed repeatedly to obtain the MAD and LSD empirical sampling distributions.

Details of each step are described in the Appendix 2.

Table 1. True and misspecified Q-matrices

	True **Q**				Misspecified* **Q**			
Item	A1	A2	A3	A4	A1	A2	A3	A4
1	1	0	0	0	1	0	0	0
2	0	1	0	0	0	1	0	0
3	0	0	1	0	0	0	1	0
4	0	0	0	1	**1**	0	0	1
5	1	1	0	0	1	1	0	0
6	1	0	1	0	1	0	1	0
7	1	0	0	1	**0**	0	0	1
8	0	1	1	0	0	1	1	0
9	0	1	0	1	0	1	0	1
10	0	0	1	1	0	0	1	1
11	0	1	1	1	0	1	1	1
12	1	0	1	1	**0**	0	1	1
13	1	1	0	1	1	1	0	1
14	1	1	1	0	1	1	1	0
15	1	1	1	1	1	1	1	1

* Changes introduced on items 4, 7, and 12 appear in boldface.

Procedure for Study II

The procedure of the second study was the same as that in Study I, but introducing misspecifications on **Q**-matrix in the last step. We randomly permutated 20% of all **Q**-matrix entries, but only the attribute A1 was manipulated (see Table 1).

The proportion of MAD and LSD values that exceeded the cut-offs was calculated in order to evaluate the sensitivity of the indices to the misspecifications introduced. Also, the means (over the 500 replications) of LSD were compared with LSD values obtained in Study I to investigate the increase due to misspecifications of the **Q**-matrix.

Design

The independent variables are: Sample size (300, 500, 1000), number of attributes required for the item (1, 2, 3, 4), and type of model employed (LSDM-C, LSDM-D). The combinations of the variable levels result in 24 design conditions. The dependent variables for both studies are MAD and LSD indices.

Table 2. True and estimated item parameters

LSDM-C								LSDM-D							
True		300		500		1000		True		300		500		1000	
a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
0.300	-2.000	0.551	-1.078	0.471	-1.245	0.391	-1.538	0.300	1.250	0.497	0.700	0.420	0.880	0.358	1.064
0.500	-1.750	0.693	-1.305	0.597	-1.494	0.526	-1.699	0.500	1.500	0.668	1.155	0.593	1.305	0.528	1.461
0.600	-1.500	0.743	-1.250	0.661	-1.414	0.617	-1.507	0.600	1.750	0.793	1.367	0.698	1.557	0.628	1.720
0.800	-1.250	0.881	-1.190	0.818	-1.247	0.802	-1.273	0.800	2.000	1.012	1.670	0.905	1.828	0.827	1.984
0.554	0.372	0.592	0.374	0.573	0.400	0.558	0.376	0.523	-0.790	0.610	-0.697	0.563	-0.764	0.533	-0.801
0.634	0.392	0.666	0.410	0.645	0.418	0.644	0.398	0.560	-0.533	0.608	-0.514	0.580	-0.561	0.562	-0.548
0.706	0.264	0.758	0.267	0.722	0.260	0.713	0.263	0.605	-0.185	0.657	-0.185	0.627	-0.187	0.613	-0.195
0.639	-0.008	0.666	0.004	0.653	-0.010	0.644	-0.002	0.582	-0.019	0.609	-0.024	0.599	-0.024	0.584	-0.020
0.858	0.021	0.895	0.013	0.866	0.024	0.865	0.028	0.744	0.318	0.787	0.314	0.765	0.325	0.754	0.315
0.962	-0.030	0.998	-0.032	0.982	-0.030	0.983	-0.030	0.894	0.496	0.917	0.503	0.931	0.481	0.894	0.503
0.949	0.713	0.990	0.709	0.966	0.730	0.966	0.717	0.858	-0.492	0.910	-0.469	0.891	-0.498	0.854	-0.506
0.758	1.025	0.833	0.983	0.775	1.046	0.758	1.040	0.720	-0.985	0.801	-0.919	0.758	-0.966	0.735	-0.994
0.696	1.166	0.759	1.106	0.730	1.155	0.707	1.182	0.586	-1.329	0.713	-1.140	0.636	-1.257	0.591	-1.355
0.765	1.126	0.840	1.071	0.787	1.138	0.774	1.139	0.714	-1.341	0.826	-1.199	0.762	-1.304	0.727	-1.351
0.799	1.541	0.923	1.408	0.852	1.505	0.807	1.566	0.686	-1.689	0.853	-1.421	0.778	-1.558	0.708	-1.683

Note: The table shows the mean of estimated parameters for each condition.

Real data application.

A mathematic test of 15 items was applied to an intentional sample of 2897 students of several Spanish schools. The cognitive structure, specified in a **Q**-matrix contains four contents (space, numbers, data and measurement) and two processes (reproduction and connections). The data matrix was calibrated using the 2PL model, and the parameters were used to execute LSDM-C. MAD and LSD values were analyzed using the cut points defined in the simulations.

An R routine was created and employed in order to execute the simulation procedures, and the real data application. The data bases were calibrated with *ltm library* for R (Rizopoulos, 2006).

RESULTS

Study 1. Description of the Distributions of MAD and LSD

LSD values. The percentiles and descriptive statistics of the LSD for both models, on 6 of the 31 fixed ability levels, appear in Table 3. There are three main results: First, independently of the sample size, under the LSDM-C the LSD values are high at low ability levels (θ = -3) and decrease as the ability increases. For example, when N = 300, the mean of LSD across the 500 replications ranges from 0.140 (at θ = -3) to 0.014 (at θ = 3). This may be explained by the conjunctive nature of the model. Contrariwise, under the LSDM-D, the LSD values are high at high ability levels. For example, when N = 300, the LSD mean range from 0.015 (at θ = -3) to 0.130 (at θ = 3). Similar tendencies can be observed for other sample sizes (see Table 3).

Second, under both models (LSDM-C and LSDM-D), the LSD values slightly decrease as the sample size increases. For example, in the case of the conjunctive model, when θ = -3, the mean of LSD is 0.140 (N = 300), 0.113 (N = 500) and 0.097 (N = 1000).

Third, also for both LSDM models, the LSD distributions converge to normal when sample size increases (see Table 3).

MAD values. Table 4 exhibits the characteristics and descriptive statistics of the MAD distributions. There are three main results: First, independently of the type of model or the sample size, the MAD values are high when only one attribute is required and decreases as more attributes are involved. For example, for the disjunctive model and N = 300, the mean of MAD values across the 500 replications ranges from 0.048 (one attribute) to 0.027 (four attributes).

Table 3. Descriptive statistics of the LSD distribution

Model	Sample	Theta	Population	Mean	*SD*	Kurtosis	Skewness	P(0.95)	P(0.99)	P(0.999)	P(0.9999)*
LSDM-C	300	-3.0	0.070	0.140	0.032	-0.064	0.377	0.198	0.223	0.248	0.253
		-2.0	0.039	0.085	0.021	0.074	0.432	0.122	0.141	0.157	0.161
		-1.0	0.017	0.042	0.011	0.802	0.670	0.062	0.075	0.086	0.086
		0.0	0.005	0.017	0.005	3.185	0.999	0.025	0.030	0.042	0.043
		1.0	0.005	0.017	0.004	0.187	0.317	0.023	0.026	0.030	0.031
		2.0	0.006	0.017	0.004	0.238	0.519	0.025	0.029	0.031	0.032
		3.0	0.006	0.014	0.004	0.297	0.544	0.020	0.025	0.027	0.027
	500	-3.0	0.070	0.113	0.025	0.425	0.261	0.154	0.173	0.204	0.208
		-2.0	0.039	0.068	0.016	0.353	0.316	0.096	0.108	0.126	0.129
		-1.0	0.017	0.033	0.009	-0.080	0.352	0.049	0.054	0.060	0.060
		0.0	0.005	0.013	0.003	0.170	0.445	0.019	0.022	0.025	0.025
		1.0	0.005	0.014	0.003	0.035	0.300	0.019	0.022	0.024	0.026
		2.0	0.006	0.015	0.004	0.026	0.269	0.021	0.024	0.026	0.027
		3.0	0.006	0.013	0.003	-0.024	0.300	0.018	0.022	0.022	0.022
	1000	-3.0	0.070	0.097	0.020	-0.054	0.197	0.133	0.145	0.160	0.161
		-2.0	0.039	0.058	0.013	-0.192	0.230	0.080	0.090	0.098	0.099
		-1.0	0.017	0.027	0.007	-0.308	0.313	0.039	0.043	0.046	0.047
		0.0	0.005	0.010	0.002	-0.260	0.338	0.015	0.016	0.017	0.017
		1.0	0.005	0.011	0.003	0.029	0.282	0.015	0.018	0.018	0.019
		2.0	0.006	0.012	0.003	-0.160	0.390	0.018	0.020	0.022	0.022
		3.0	0.006	0.011	0.003	-0.053	0.445	0.016	0.018	0.020	0.020
LSDM-D	300	-3.0	0.007	0.015	0.004	-0.080	0.201	0.021	0.024	0.025	0.025
		-2.0	0.007	0.017	0.004	-0.162	0.300	0.024	0.027	0.028	0.028
		-1.0	0.005	0.016	0.004	0.067	0.306	0.022	0.025	0.028	0.028
		0.0	0.005	0.016	0.004	0.367	0.534	0.024	0.027	0.032	0.034
		1.0	0.015	0.037	0.010	1.933	0.837	0.053	0.064	0.081	0.087
		2.0	0.034	0.076	0.018	0.863	0.597	0.108	0.125	0.148	0.151
		3.0	0.061	0.130	0.028	0.486	0.485	0.177	0.204	0.229	0.235
	500	-3.0	0.007	0.014	0.003	0.155	0.350	0.020	0.023	0.025	0.026
		-2.0	0.007	0.015	0.004	0.208	0.351	0.022	0.025	0.027	0.027
		-1.0	0.005	0.013	0.003	-0.065	0.194	0.019	0.021	0.023	0.024
		0.0	0.005	0.013	0.003	0.156	0.368	0.018	0.021	0.023	0.024
		1.0	0.015	0.030	0.008	0.342	0.212	0.042	0.047	0.060	0.061
		2.0	0.034	0.062	0.015	0.087	0.184	0.085	0.096	0.115	0.117
		3.0	0.061	0.107	0.023	0.002	0.151	0.143	0.162	0.178	0.178
	1000	-3.0	0.007	0.013	0.003	0.128	0.381	0.017	0.020	0.024	0.025
		-2.0	0.007	0.013	0.003	-0.098	0.278	0.019	0.022	0.023	0.023
		-1.0	0.005	0.011	0.003	-0.047	0.192	0.015	0.017	0.019	0.019
		0.0	0.005	0.010	0.002	-0.284	0.171	0.014	0.015	0.016	0.016
		1.0	0.015	0.024	0.006	-0.224	0.421	0.036	0.040	0.043	0.043
		2.0	0.034	0.051	0.012	-0.200	0.290	0.072	0.081	0.085	0.085
		3.0	0.061	0.088	0.019	-0.216	0.174	0.121	0.134	0.142	0.143

* Values selected as cut point for study II

Second, in both models, the MAD values slightly decrease as the sample size increases. For example, for the case of the disjunctive model and one attribute, the mean of MAD values are: 0.049 (N = 300), 0.039 (N = 500) and 0.030 (N = 1000). This occurs independently of the number of attributes required by the item (see Table 4).

Third, the MAD distributions presents positive skew ($g_1 > 0.5$) and are "peaked" ($g_2 > 0.5$).

Table 4. Descriptive statistics of the MAD distribution

Model	Sample	Attributes	Population	Mean	*SD*	Kurtosis	Skewness	P(0.95)	P(0.99)	P(0.999)	P(0.9999)*
LSDM-C	300	1	0.014	0.048	0.025	1.755	1.157	0.093	0.125	0.157	0.166
		2	0.016	0.036	0.020	0.615	0.891	0.074	0.091	0.110	0.115
		3	0.011	0.028	0.016	0.305	0.783	0.059	0.071	0.083	0.085
		4	0.012	0.029	0.015	2.208	1.136	0.057	0.073	0.098	0.106
	500	1	0.014	0.038	0.019	1.230	1.047	0.074	0.091	0.113	0.119
		2	0.016	0.030	0.016	0.674	0.867	0.060	0.078	0.092	0.095
		3	0.011	0.025	0.014	0.592	0.901	0.052	0.065	0.076	0.078
		4	0.012	0.025	0.012	-0.092	0.740	0.047	0.056	0.067	0.071
	1000	1	0.014	0.029	0.014	2.103	1.339	0.057	0.075	0.090	0.093
		2	0.016	0.025	0.014	1.186	0.998	0.049	0.064	0.080	0.083
		3	0.011	0.020	0.012	0.463	0.914	0.042	0.054	0.061	0.062
		4	0.012	0.022	0.011	0.799	1.031	0.044	0.057	0.065	0.067
LSDM-D	300	1	0.016	0.049	0.025	1.126	1.081	0.097	0.124	0.144	0.148
		2	0.015	0.036	0.020	0.534	0.803	0.072	0.090	0.108	0.112
		3	0.012	0.029	0.015	0.979	0.884	0.057	0.071	0.088	0.091
		4	0.016	0.027	0.013	1.883	1.198	0.052	0.068	0.086	0.090
	500	1	0.016	0.039	0.020	1.561	1.175	0.078	0.097	0.124	0.129
		2	0.015	0.030	0.017	0.635	0.904	0.063	0.079	0.093	0.097
		3	0.012	0.025	0.014	0.618	0.878	0.053	0.064	0.075	0.076
		4	0.016	0.026	0.011	-0.260	0.419	0.046	0.051	0.067	0.072
	1000	1	0.016	0.030	0.015	2.400	1.352	0.060	0.077	0.098	0.102
		2	0.015	0.025	0.014	1.238	1.032	0.050	0.066	0.080	0.084
		3	0.012	0.021	0.012	1.003	0.979	0.045	0.056	0.067	0.068
		4	0.016	0.024	0.012	0.218	0.860	0.049	0.059	0.064	0.064

* Values selected as cut point for study II

The percentiles of the MAD distributions are presented in Figure 1. As can be seen, independently of the type of LSDM model, the highest MAD values correspond to the sample of 300 and , conversely, the lowest MAD values correspond to the sample of 1000. Figure 1 also shows that the distributions of MAD are affected by the number of attributes involved.

Study 2. Detection of Q-matrix Misspecifications

LSD values. Figure 2 present the means of the LSD values before and after introducing misspecifications in the **conjunctive** model and with a sample of 300. As expected, the LSD values increase after introducing misspecifications, especially at low and medium ability levels. The same tendencies occur with the other sample sizes (500 and 1000). As can also be seen in Figure 2, there are discrepancies between original and true **Q** matrices, which indicates that the LSD index is affected by the IRT calibration process. Other result shows that the means of LSD slightly decrease with the sample size. For example, for $\theta = -3$, the LSD mean is 0.14 in the sample of $N = 500$ and 0.12 in the sample of $N = 1000$.

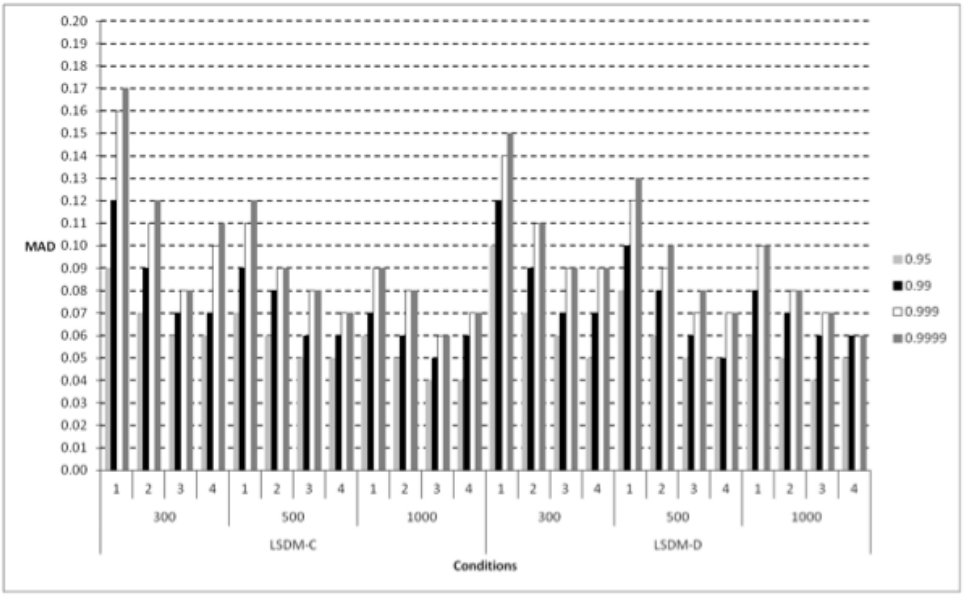


Figure 1. Percentiles 95, 99, 999 and 9999 of the MAD distribution by sample size and number of attributes.

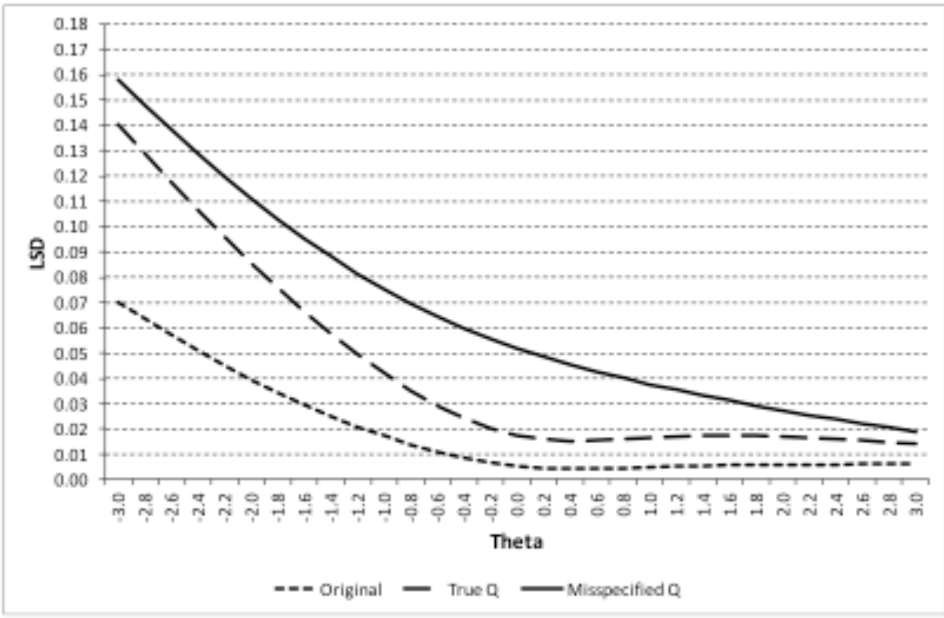


Figure 2. Mean of LSD N = 300 (LSDM-C)

Figure 3 presents the means of the LSD values before and after introducing misspecifications in the **disjunctive** model with a sample of 300. Again, the LSD values increase after introducing misspecifications but the comparison of Figures 2 and 3 shows that the trend in the two models is opposite. Specifically, under the disjunctive model, the LSD values increase as ability increases, whereas in under the conjunctive model, the LSD values decrease as ability increases. Results also show that, independently of the model, the means of LSD values slightly decrease with the sample size. For example, for θ = 3 the LSD mean is 0.14 in the sample of *N* = 500 and 0.13 in the sample of *N* = 1000.

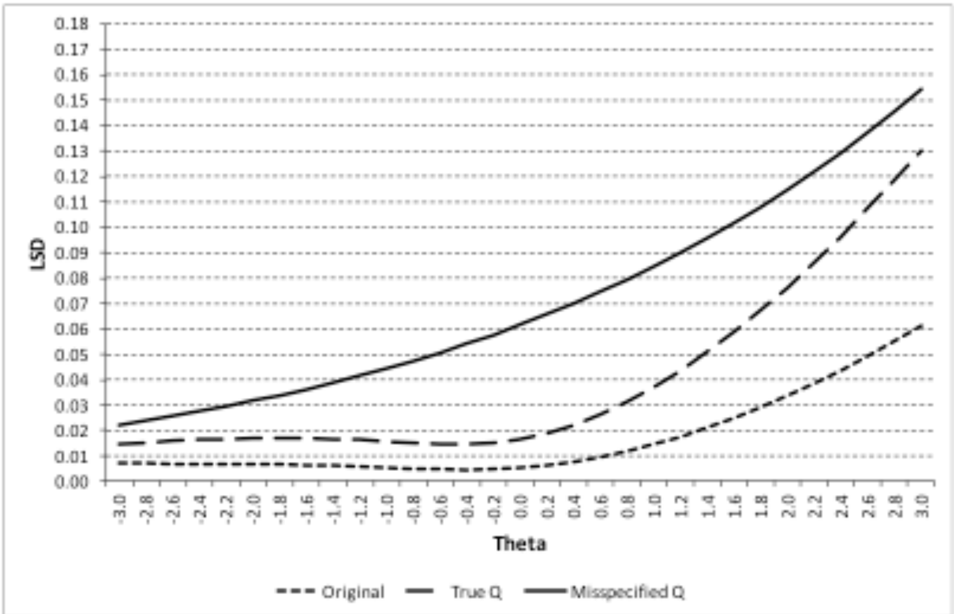


Figure 3. Mean of LSD N = 300 (LSDM-D)

MAD values. Figure 4 presents the proportion (in the 500 replications) of MAD values that exceed the cut points in under both LSDM models. In the sample of *N* =1000, all misspecified items (4, 7, and 12) are detected through the MAD values, because the MADs corresponding to these items exceed the cut-offs of this condition in all the replications. However, under both models, 38% of the MAD values of the item 1, which was not misspecified, also exceed the established cut-off. The high proportion of erroneous detection of the item 1 may be explained by the fact that this item is extremely easy and has low discrimination. Under the disjunctive model, an erroneous detection occurred for the item 10 in 30% of the replications.

In the sample of N = 500, all misspecified items (1, 7 and 12) are detected in the 100% of the cases, except item 12 that is detected in 70% of the cases using the conjunctive model. Likewise, when N = 300, misspecifications in items 4 and 12 were detected only in 62% of the cases using the LSDM-C, thus indicating less accuracy in signaling misspecifcation under the conjunctive model when the sample size is small. Under the LSDM-D, however, all the misspecified items are detected, even when the sample size is small.

Applications in a mathematic test. In order to provide practical value to the simulation results, a real data example of validation under LSDM is presented now. The test is composed of 15 items of mathematics with a format of multiple choice. The test was applied to an intentional sample of 2897 students of fourth grade in several Spanish schools. The data matrix was calibrated using the 2PL model. Table 5 presents the **Q**-matrix, the IRT item parameters, standard errors of parameters, MAD values and LSD values.

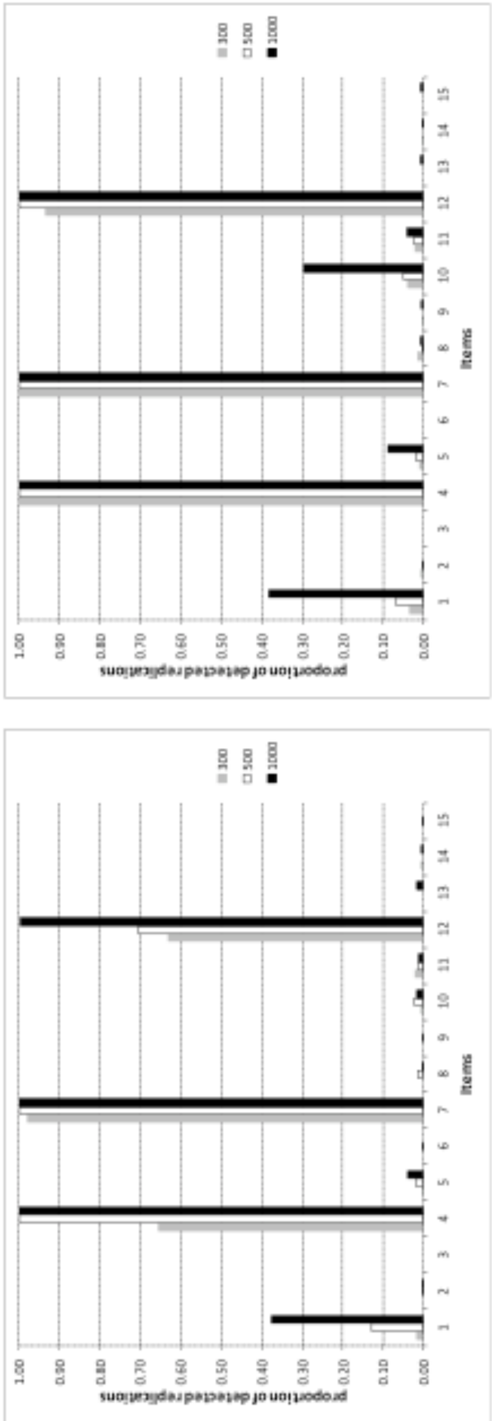
The LSDM criteria for MAD and LSD were studied according to the referred cut points for **N** = 1000, because is the more similar condition. As can be seen in Table 5, two of the fifteen items exceeds the cut-off proposed for the MAD index. These two items have in common the content of numbers (A2) but require different processes: connections (A6) for item 2 and reproduction (A5) for item 7. A re-specification of these items seems necessary, possibly of attributes related to processes.

Regarding to the LSD index, as found in other applications, the conjunctive LSDM is less accurate at low levels of the trait measured by the test, exceeding the cut-off when $\theta \leq 1$.

DISCUSSION

A correct specification of the **Q**-matrix is an important part of the design of the cognitive diagnosis procedures. For this reason, developing validation criteria that indicate possible misspecifications for items in the **Q**-matrix is of critical importance for successful application of the CDMs.

This paper presents simulation results about the empirical sample distributions and cut-offs of the LSDM indices MAD and LSD under several conditions that can be useful in applied context. Although this



LSDM-D

LSDM-C

Figure 4. Proportion of items that exceed the 9999th percentile of the MAD distribution

simulation study was complex, it could not be exhaustive with respect to all relevant assessment design and misspecification conditions, and for that reason this research constitute a first step for the knowledge of the LSDM indices behavior.

Table 5. Q-matrix, 2PL calibration and fit and MAD indices of the mathematic test

Q-matrix							IRT				LSDM-C		
Item	A_1	A_2	A_3	A_4	A_5	A_6	*a*	s.e	*B*	s.e	MAD	Theta	LSD
1	0	0	1	0	1	0	0.60	0.05	0.56	0.08	0.01	-3	**0.17**
2	**0**	**1**	**0**	**0**	**0**	**1**	1.47	0.08	-0.38	0.04	**0.11**	-2.5	**0.14**
3	1	0	0	0	0	1	1.00	0.06	-0.99	0.06	0.04	-2	**0.10**
4	1	0	0	0	0	1	1.46	0.08	-0.66	0.04	0.03	-1.5	**0.07**
5	1	0	0	0	1	0	1.15	0.07	-0.80	0.05	0.03	-1	**0.05**
6	0	0	1	0	0	1	0.66	0.05	0.45	0.07	0.05	-0.5	**0.04**
7	**0**	**1**	**0**	**0**	**1**	**0**	0.55	0.05	-0.78	0.09	**0.11**	0	**0.03**
8	0	1	0	0	0	1	0.87	0.05	0.44	0.06	0.02	0.5	**0.03**
9	0	1	0	0	1	0	0.75	0.05	0.16	0.06	0.03	1	**0.02**
10	0	0	0	1	0	1	0.60	0.05	0.16	0.07	0.04	1.5	0.02
11	0	0	1	0	0	1	0.64	0.05	1.22	0.11	0.04	2	0.02
12	0	1	0	0	1	0	1.22	0.07	0.21	0.04	0.06	2.5	0.02
13	0	1	0	0	0	1	1.17	0.07	0.55	0.05	0.06	3	0.02
14	0	1	0	0	0	1	0.58	0.05	0.46	0.08	0.08		
15	0	0	0	1	0	1	0.40	0.05	1.09	0.15	0.04		

Additionally, results indicate that the behavior of the indices depends of the design of the assessment and the indexes do not allow cut-offs fixed for all situations, so in practice the researchers interested in the application of LSDM for **Q**-matrix validation using the cutoffs provided here, have to look the simulation condition more similar to their assessment design, as was exemplified in the empirical study presented here. Also, the present study design could be extended to cover a broader range of assessment design conditions.

The purpose of this research was to obtain information about the empirical sample distribution of LSD and MAD values and to identify cut-off values for correctly specified items that permit to refine the validation process proposed by Dimitrov (2007). This objective has been met and the results provided can be useful to practitioners during the process of validation and re-specification of the **Q**-matrix, because both indices (LSD and MAD) are complementary and may be employed in applied evaluations in two phases, namely (a) the LSD values can be used as indicators of overall adjustment of the **Q**-matrix at specific ability levels and then (b) the MAD values can be used as indicators of items that should be re-specified.

The results indicate that the MAD values are affected by the number of attributes required by the item. It is therefore difficult to find a unique cut point to evaluate the validity of the **Q**-matrix based on the ICC recovery for items. It seems more reasonable to propose several cut points according to the number of attributes involved.

Following the results of the Study II and the real data application, we can conclude that the MAD is a useful statistic to detect items with misspecifications. Additionally, the empirical cut points proposed herein, based on the number of attributes required by the item, seem more accurate than the rules-of-thumb proposed by Dimitrov (2007). However, a deeper study of the sensitivity and adequacy of the proposed cut points, with different degrees of **Q**-matrix misspecification, is necessary. Also, more applications using the cut points should be made in different cognitive structures of educational and psychological tests.

Regarding the LSD index, it may be of interest to develop a statistical fit index based on the area between the item characteristic curves produced when the **Q** is true and when **Q** is misspecified, with the purpose of such statistic to minimize the discrepancy between the curves caused by IRT calibration and to assess the fit of the LSDM model at various ability levels.

As the present work represents a first effort to the study of LSDM validation indices, much work remains to be done in this area. It is necessary to study misspecifications in the **Q**-matrix under more conditions and to analyze the behavior of the LSD, MAD (and possibly other) criteria across a wide range of situations. Also, the data could be calibrated with using other IRT models, such as the Rasch model and 3PL model. It may also be of interest to study the effects of item misfit to with the IRT model on the power of LSDM-related criteria for detecting **Q**-misspecifications.

In several practical applications of the LSDM we have found that the simple inspection of both indices may be very useful, specially when have alternative **Q**-matrices to test. In real applications, as the presented here, the "true" **Q**-matrix is unknown but a comparison of alternative **Q**s from different cognitive models may be used and both indices can be compared for the competing models: lower values of MAD and LSD support model choice decisions and this procedure may be used in conjunction with any CDM.

RESUMEN

Detección de errores de especificación en la matriz Q utilizando dos criterios de validación de estructuras cognitivas con el Modelo de las Distancias Mínimo Cuadráticas (LSDM). Los Modelos de Diagnóstico Cognitivo (MDC) tienen por objeto proporcionar información sobre el grado en que los individuos dominan atributos específicos para resolver correctamente los items de un test. La matriz **Q** es un elemento clave en la aplicación de los MDC porque contiene vínculos entre items y atributos que representan la estructura cognitiva propuesta para resolver la prueba. Por medio de un estudio de simulación, se determinó el rendimiento de dos estadísticos de ajuste (MAD y LSD) para detectar errores de especificación en la matriz **Q** dentro del marco del modelo de la distancia mínimo cuadrática. Los factores manipulados en el diseño del test incluyen: número de encuestados (300, 500, 1000), número de atributos (1, 2, 3, 4), y el tipo de modelo (conjuntivo vs disyuntivo). Se investigó el comportamiento de los valores MAD y LSD bajo una correcta especificación de **Q**, con errores de especificación en **Q** y en una aplicación de datos reales. Los resultados muestran que los dos índices son sensibles a los errores de especificación de **Q**, por este motivo se proponen puntos de corte para usar en aplicaciones del modelo.

REFERENCES

Corter, J. E. (1995). Using clustering methods to explore the structure of diagnostic tests. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 305-326). Hillsdale, NJ: Erlbaum.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*, 343-362. doi:10.1111/j.1745-3984.2008.00069

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130. doi: 10.3102/1076998607309474

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69,* 333-353. doi: 10.1007/BF02295640

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Erlbaum.

Dimitrov, D. M. (2007). Least squares distance method of cognitive validation and analysis for binary items using their item response theory parameters. *Applied Psychological Measurement, 31*, 367-387. doi: 10.1177/0146621606295199

Dimitrov, D. M., & Atanasov, D. V. (2011). Conjunctive and disjunctive extensions of the least squares distance model of cognitive diagnosis. *Educational and Psychological Measurement*. doi: 10.1177/0013164411402324

Dimitrov, D. M., & Raykov, T. (2003). Validation of cognitive structures: A structural equation modeling approach. *Multivariate Behavioral* Research, *38(1)*, 1-23. doi: 10.1207/S15327906MBR3801_1

Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika, 49*(2), 175-186. doi: 10.1007/BF02294171

Embretson, S. (1993). Psychometric models for learning and cognitive processes. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 125-150). Hillsdale, NJ: Erlbaum.

Fischer, G. H. (1995). The linear logistic test model. In G. H. Fisher & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications (*pp. 131-156). New York: Springer-Verlag.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement, 29*, 262-277. doi: 10.1177/0146621604272623

Henson, R., Templin, J. & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74(2)*, 191-210. doi: 10.1007/s11336-008-9089-5

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*, 258-272. doi: 10.1177/01466210122032064

Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-Hall.

Medina-Diaz, M. (1993). Analysis of cognitive structure using the linear logistic test model and quadratic assignment. *Applied Psychological Measurement*, *17(2)*, 117-130. doi: 10.1177/014662169301700202

Rao, C. R. & Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. New York: Wiley.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, *17*(5).

Romero, S. J. (2010). Propiedades y aplicaciones del método de las distancias mínimo cuadráticas (LSDM) para la validación y análisis de atributos cognitivos [Properties and applications of the least squares distance method (LSDM) for validation and analysis of cognitive attributes]. (Doctoral Dissertation in Spanish, Universidad Autónoma de Madrid). Retrieved at: http://digitool-uam.greendata.es:1801/webclient/DeliveryManager?pid=30756

Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68(1)*, 78-96. doi: 10.1177/0013164407301545

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics, 10(1)*, 55-73. doi: 10.2307/1164930

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327-359). Hillsdale, NJ: Erlbaum.

Templin, J. L. & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287-305. doi: 10.1037/1082-989X.11.3.287

von Davier, M. (2010). Hierarchical mixtures of diagnostic models. *Psychological Test and Assessment Modeling, 52(1)*, 8-28.

APPENDIX 1

Some theoretical properties of the LSD

Let Z be the probability of the item response model (i.e. the 2PL model) and Y be the probability reproduced by the cognitive model (computed by either equation (1) or (2)). The LSD estimate is found by minimizing the squared difference

$$\begin{aligned} SD_1 &= \sum_i (\log y_i - \log z_i)^2 \\ &= \sum_i \left(\log \frac{y_i}{z_i} \right)^2, \end{aligned} \qquad \text{(A1)}$$

where the subscript i runs over all items and values of θ. Suppose that h is the difference between the cognitive model minus the IRT one ($h_i = y_i - z_i$). Then

$$\begin{aligned} SD_1 &= \sum_i \left(\log \frac{z_i + h_i}{z_i} \right)^2 \\ &= \sum_i \left(\log \left(1 + \frac{h_i}{z_i} \right) \right)^2. \end{aligned} \qquad \text{(A2)}$$

The term h_i / z_i increases when z_i decreases. This has an important implication for estimation: the difference h_i receives more weight as the probability of the IRT model decreases. That is, the estimation algorithm tries harder to fit the two item characteristic curves at the lower end of the ability continuum. On the other hand, differences between the IRT and the cognitive models at medium or high values of θ are deemed less important, and have a less remarkable effect on the estimates of the cognitive model. One way to overcome this unwilling property would be to minimize the function

$$SD_2 = \sum_i (y_i - z_i)^2, \qquad \text{(A3)}$$

subject to the constraint $0 \le y_i \le 1$. The function SD_2 gives the same weight to a given difference irrespective of the point of θ where it occurs. Differences can also be weighted by the density of θ, so that a difference would have more weight if it occurs at a point of θ of high density. In that case, estimation would be based on the function:

$$SD_3 = \sum_i f(\theta_i)(y_i - z_i)^2. \qquad \text{(A4)}$$

The investigation of deviation functions SD_2 and SD_3 is deferred to future research.

Regarding the function SD_1, finding the LSD estimate is the same as finding the least squares estimate of the linear equation $\boldsymbol{l} = \boldsymbol{Qx}$ subject to the constraint $\boldsymbol{x} \leq \boldsymbol{0}$. The solution to this equation depends on the rank of Q. First, the equation is consistent if and only if $\boldsymbol{Q}^{-}\boldsymbol{Ql} = \boldsymbol{l}$ for some generalized inverse, $\boldsymbol{Q}^{-}$, of Q. Consistency implies that LSD = 0 although we do not expect that this holds in general.

When the equation is not consistent, the general form of the least squares solution to the linear equation is (Rao & Mitra, 1971):

$$\boldsymbol{x}_s = \boldsymbol{Q}^{-}\boldsymbol{l} + (\boldsymbol{I} - \boldsymbol{Q}^{-}\boldsymbol{Q})\boldsymbol{v}, \quad \text{(A5)}$$

where $\boldsymbol{v}$ is an arbitrary vector, I is an identity matrix and the constraint $\boldsymbol{x} \leq \boldsymbol{0}$ has not been taken into account. When Q has full column rank, $\boldsymbol{Q}^{-}$ is a left inverse: $\boldsymbol{Q}^{-}=(\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{Q}'$. Then $\boldsymbol{x}_s = \boldsymbol{Q}^{-}\boldsymbol{l} + (\boldsymbol{I} - \boldsymbol{Q}^{-}\boldsymbol{Q})\boldsymbol{v} = \boldsymbol{Q}^{-}\boldsymbol{l} + (\boldsymbol{I} - \boldsymbol{I})\boldsymbol{v} = \boldsymbol{Q}^{-}\boldsymbol{l}$ and $\boldsymbol{x}_s$ is unique. Moreover, equation (A5) implies that there are infinite least squares solutions (values of $\boldsymbol{x}_s$) to the equation when Q is deficient in rank. In substantive terms, there are infinite different vector of parameters for the cognitive models that minimize the LSD. Thus, interpretation of the cognitive model is arbitrary. For these reasons, full column rank of Q must be a minimum requirement for the cognitive model.

APPENDIX 2

Details of the simulation procedure

Step 1. An arbitrary **Q** matrix with at least one of four attributes and fifteen items was generated by including all the possible binary combinations of 0-1, so that the items 1 to 4 require one attribute; items 5 to 10 require two attributes, items 11 to 14 require three attributes and item15 requires the four attributes.

Step 2. The probabilities of correct performance on attributes A_k, $P(A_k = 1|\theta_i)$, were generated for 31 equally spaced ability levels between -3 and 3 (on the logit scale) using the 2PL model in IRT. The discrimination and difficulty parameters of attribute probability curves (APCs) were selected to obtain items with medium discriminations and varying difficulty levels. The parameters of the APCs for the LSDM-C were: A1 (a = 0.30, b = -2), A2 (a = 0.50, b = -1.75), A3 (a = 0.60, b = -1.5), and A4 (a = 0.80, b = -1.25). For the LSDM-D, we used the same "*a*" parameters but opposite "*b*" parameters (b_1 = 2, b_2 = 1.75, b_3 = 1.5, and b_4 = 1.25) in order to compensate for the extremely easy items generated by the disjunctive nature of the model.

Step 3. The values of P_{ijLSDM} were computed from the APCs using the LSDM-C and LSDM-D, (equations 1 and 2, respectively) for each of the 31 ability levels.

Step 4. For each of the 15 items, the couple of 2PL parameters (a, b) that minimize the difference between P_{ij} and P_{ijLSDM} was found. This was done by looking examining couples of 2PL parameters that best fit the values of P_{ijLSDM} by through the use of an iterative procedure.

Step 5. The LSDM application using the parameters found in step 4 allows us to obtain population values of MAD and LSD that appear in columns "population" of Tables 3 and 4. As can be seen, the MAD and LSD for the iterative procedure are not exactly, yet close to, zero.

Step 6. Simulated were the responses of 300, 500, and 1000 examinees for the 15 items using the original parameters presented in Table 2. The 2PL was estimated from the simulated responses and the cognitive models (LSDM-C and LSDM-D) were fitted to the estimated (a, b) using the same Q-matrix as in step 2. Finally, the MAD and LSD were computed for each item across all fixed ability levels.

Step 7. The fifth step was replicated 500 times to obtain the distribution of the discrepancy measure for each condition.

(Manuscript received: 4 September 2012; accepted: 24 May 2013)