

Análisis de la estructura cognitiva de la competencia científica en PISA 2006 mediante el método de las distancias mínimo-cuadráticas: el caso español

Sonia J. Romero, Xavier G. Ordóñez*, Esther López Martín* y Enrique Navarro*
Universidad Autónoma de Madrid y * Universidad Complutense de Madrid

Los modelos de diagnóstico cognitivo integran la Psicometría y la Psicología Cognitiva con el objetivo de realizar mediciones detalladas de los procesos o atributos cognitivos requeridos para resolver los ítems de un test. Dentro de este contexto, el método de las distancias mínimo-cuadráticas (en inglés, LSDM) utiliza los parámetros de los ítems, estimados con alguno de los modelos de la Teoría de Respuesta al Ítem (TRI), para analizar los atributos y proporcionar evidencia sobre la validez de la estructura cognitiva. La Organización para la Cooperación y el Desarrollo Económico (OCDE) define tres procesos cognitivos subyacentes a la ejecución de los ítems de ciencias en PISA-2006: a) identificar cuestiones científicas, b) explicar fenómenos científicamente, y c) utilizar pruebas científicas. El objetivo del presente trabajo es analizar estos tres procesos y acopiar evidencia sobre la validez de la estructura propuesta por la OCDE, utilizando para ello el LSDM. Un segundo objetivo es comparar las comunidades autónomas participantes en cuanto al dominio de los tres atributos. Los resultados evidencian que la estructura propuesta explica adecuadamente la ejecución de los estudiantes en los ítems y que existen diferencias significativas entre tres comunidades autónomas en cuanto al dominio de los mismos.

Analysis of the cognitive structure of scientific competence in PISA-2006 by means of the Least Squares Distance Method (LSDM): The Spanish case. The cognitive diagnosis approach links Psychometrics and Cognitive Psychology in order to obtain detailed measurements of the processes and cognitive attributes required to solve test items. In this context, the Least Squares Distance Method (LSDM) employs item parameters, calibrated with any Item Response Theory (IRT) model, to analyze the attributes and provide evidence about the validity of the cognitive structure. The Organization for Economic Cooperation and Development (OECD) determines three cognitive attributes underlying students' performance in the PISA-2006 science items: a) identifying scientific issues, b) explaining phenomena scientifically and c) using scientific evidence. The aim of the present work is to analyze these processes and collect evidence of the cognitive structure proposed by the OECD, using the LSDM. A second aim is to compare the probabilities of mastering the attributes of the Spanish Communities. The results show that the proposed structure is appropriate to explain students' proficiency in the science items, and also, significant differences between three Spanish Communities were found.

En las últimas décadas ha ocurrido un interés considerable en el desarrollo de modelos psicométricos que incorporan componentes (procesos cognitivos, estrategias, operaciones, reglas o atributos) que subyacen a la ejecución de los ítems de los test. Algunos modelos como el LLTM (Linear Logistic Test Model; Fisher, 1973), el MLTM (Multicomponent Latent Trait Model; Whitely, 1980), el GLTM (General Latent Trait Model; Embretson, 1984) y

el GDM (General Diagnostic Model; von Davier y Yamamoto, 2004) buscan integrar la Teoría de Respuesta al Ítem (TRI) con la teoría cognitiva mediante la descomposición de parámetros de los ítems o de la habilidad de los sujetos en componentes cognitivos.

Otro grupo de modelos, conocidos como modelos de diagnóstico cognitivo, se centran en la clasificación de los examinados de acuerdo al dominio de los atributos. Algunos ejemplos de este tipo de modelos son el RSM (Rule Space Model; Tatsuoka, 1985, 1995), los modelos DINA y NIDA (Junker y Sijtsma, 2001; Henson y Douglas, 2005), el UM (Unified Model; Dibello, Stout y Roussos, 1995), el FM (Fusion Model; Hartz, Roussos y Stout, 2002) y la aproximación por redes bayesianas (Yan, Almond y Mislavy, 2004).

Un elemento importante y común entre todos los modelos anteriormente mencionados es el uso de una matriz de especificación

en la que se expresa la estructura cognitiva del test. Dicha matriz, denominada **Q**, suele ser binaria: sus elementos asumen un valor de 1 cuando el ítem requiere el atributo y 0 en caso contrario. En algunas aplicaciones (Fischer, 1973), así como en el presente trabajo, se ha utilizado una matriz **Q** no binaria cuyos elementos son pesos o frecuencias de uso de los atributos. La matriz **Q** constituye un vínculo entre ítems y atributos, parte de concepciones teóricas, debe ser especificada con anterioridad a la aplicación de la prueba e idealmente debe ser utilizada en la fase de diseño y construcción del test (Embretson y Reise, 2000).

Cuando se emplean este tipo de modelos es fundamental acopiar evidencia sobre la validez de la estructura cognitiva expresada en la matriz **Q**. En concreto, es necesario comprobar que los atributos especificados sean utilizados por los examinados al resolver el test y estudiar si la relación propuesta entre atributos e ítems es correcta. Una mala especificación de **Q** podría conducir a inferencias equivocadas sobre el desempeño de los sujetos en cada uno de los atributos y, por tanto, a un diagnóstico erróneo de sus fortalezas y dificultades.

Se han propuesto varias metodologías para estudiar la validez de la estructura cognitiva, entre ellas se puede contar el ajuste del modelo utilizado (Whitely y Schneider, 1981), la regresión lineal (Whitely, 1980), los estudios correlacionales (Embretson, 1984), el estudio de las tasas de clasificación correcta (Tatsuoka, 1985), la inclusión de parámetros que cuantifican el error causado por una mala especificación de **Q** (Dibello, Stout y Roussos, 1995; Junker y Sijtsma, 2001), el método de asignación cuadrática (Medina-Díaz, 1993) y el modelamiento por ecuaciones estructurales (Dimitrov y Raykov, 2003; Romero, Ponsoda y Ximénez, 2006).

Todas estas propuestas requieren información sobre las puntuaciones de los sujetos en cada uno de los ítems e incluso algunas de ellas necesitan conocer el desempeño de los sujetos en cada uno de los atributos. Un procedimiento desarrollado por Dimitrov (2007), llamado método de las distancias mínimo-cuadráticas (LSDM), no necesita dicha información. En su lugar, el LSDM precisa de las estimaciones de los parámetros de los ítems obtenidas mediante algún modelo de la TRI (Rasch, modelos logísticos de 1, 2 o 3 parámetros) para estimar las probabilidades de dominio de los atributos. Esto permite al investigador utilizar el LSDM aunque no disponga de los datos de las respuestas de los sujetos.

El LSDM supone que la probabilidad de respuesta correcta a un ítem es igual a la probabilidad de que todos los atributos requeridos por dicho ítem sean correctamente aplicados, esto es:

$$P_{ij} = \prod_{k=1}^K [P(A_k = 1 | \theta_i)]^{q_{jk}} \tag{1}$$

Donde:

P_{ij} es la probabilidad de respuesta correcta en el ítem j para una persona i en el nivel de habilidad θ_i .

$P(A_k = 1 | \theta_i)$ es la probabilidad de ejecutar correctamente el atributo A_k para un examinado en el nivel de habilidad θ_i .

q_{jk} es el elemento de la matriz **Q** que une el ítem j con el atributo A_k .

Tomando el logaritmo natural en ambos lados de la ecuación (1) se llega al sistema de ecuaciones $\mathbf{L} = \mathbf{QX}$, donde **L** es un vector (conocido) de elementos $\ln(P_{ij})$; **Q** es la matriz de especificación (conocida) y **X** es el vector de incógnitas $X_k = \ln(P[A_k = 1 | \theta_i])$. Las soluciones para el vector **X** buscan minimizar la norma eucli-

diana de $\|\mathbf{QX-L}\|$ utilizando el método de las distancias mínimo-cuadráticas. Una vez se ha encontrado la solución para **X**, se tiene que $P(A_k = 1 | \theta_i) = \exp(X_k)$. Como resultado se obtiene la curva de probabilidad para cada atributo a través de niveles de habilidad fijos en la escala logística.

Dimitrov (2007) propone los siguientes criterios heurísticos para la validación de estructuras cognitivas a través del LSDM:

1. El comportamiento lógico y monótonico de las curvas de probabilidad de los atributos. La monotonidad implica que a medida que aumenta la habilidad de las personas, aumenta también la probabilidad dominio de los atributos. La falta de monotonidad evidencia algún fallo en la especificación de dicho atributo. El comportamiento lógico se refiere a que la dificultad relativa de los atributos sea consistente con la teoría subyacente que se ha usado para definirlos.
2. La recuperación de las curvas características de los ítems (CCIs), mediante el LSDM: permite conocer si cada ítem es bien explicado por el producto de las probabilidades de dominio de los atributos. La calidad de recuperación es medida por la diferencia media absoluta (MAD) entre la CCI obtenida por el modelo TRI y la CCI recuperada por el LSDM. Dimitrov (2007) propone una regla de clasificación para la recuperación: a) muy buena ($0 \leq MAD < 0.02$), b) buena ($0.02 \leq MAD < 0.05$), c) algo buena ($0.05 \leq MAD < 0.1$), d) algo pobre ($0.1 \leq MAD < 0.15$) y f) muy pobre ($MAD \geq 0.2$). $MAD = 0$ indicaría una perfecta recuperación de la CCI para el ítem en cuestión.

La competencia científica en PISA-2006

El programa PISA de la OCDE es una evaluación internacional estandarizada aplicada a alumnos de 15 años, tiene lugar cada 3 años, ha comenzado en el año 2000 y se extenderá hasta 2015. En PISA-2006 han participado 56 países y por término medio en cada país solucionaron la prueba entre 4.500 y 10.000 estudiantes (OECD, 2007). PISA-2006 abarca las áreas de competencia lectora, matemática y científica. En 2006 el hincapié se ha realizado en la competencia científica que es definida como la capacidad de aplicar conocimientos científicos para identificar problemas, adquirir nuevos conocimientos, explicar fenómenos y extraer conclusiones basadas en pruebas (OECD, 2007; y MEC, 2007). Según la OCDE (2007, 2009), la competencia científica se descompone en tres atributos o procesos cognitivos, que se han utilizado en este trabajo para definir la matriz **Q**: identificar cuestiones científicas (A_1), explicar fenómenos científicamente (A_2) y utilizar pruebas científicas (A_3).

Descripción de la prueba y especificación de **Q**

En PISA-2006 se han incluido ítems de rendimiento y de actitud. Las preguntas de la prueba de rendimiento son de formatos diversos: de Elección Múltiple simple (E.M) y Compleja (E.M.C.), de Respuesta Abierta (R.A) y de Respuesta Elaborada (R.E) con criterios objetivos de puntuación (véase tabla 1). En las preguntas sobre actitudes se utilizan escalas tipo Likert. Los ítems se agrupan de acuerdo a un estímulo común (texto y figura). Por ejemplo, del estímulo llamado *lluvia ácida* se derivan cinco preguntas, tres de las cuales se refieren a los procesos definidos por la OCDE y dos a las actitudes de los examinados.

Para realizar los análisis del presente estudio se han utilizado los 53 ítems de rendimiento aplicados a la muestra española en el cuadernillo 1. Se ha usado dicho cuadernillo porque es el que tiene mayor cantidad de ítems de ciencia y además todos los ítems que incluye son de esta área. Sin embargo, no son tenidos en cuenta los ítems que se diseñaron para estudiantes con necesidades educativas especiales. Los ítems de rendimiento en PISA han sido diseñados de manera que cada ítem requiere un solo proceso para ser resuelto, por este motivo se ha decidido especificar la matriz **Q** de acuerdo a los requerimientos cognitivos de cada uno de los 18 ítem-estímulo (o compuestos). Por ejemplo, el estímulo de efecto invernadero se compone de tres preguntas: dos de ellas requieren que el examinado use evidencia científica (A_2) y una que identifique cuestiones científicas (A_3). Los elementos de la matriz **Q** son frecuencias pues el examinado debe aplicar el atributo A_2 dos veces y el A_3 una vez para resolver correctamente este ítem compuesto. La matriz **Q** y las descripciones de los ítems se presentan en la tabla 1.

Mediante el presente trabajo se pretende acopiar evidencia sobre la calidad de la estructura cognitiva de los ítem-estímulo que miden la competencia científica en PISA-2006 y analizar los tres procesos propuestos por la OCDE para este tipo de tareas. El segundo objetivo es estudiar las diferencias entre las comunidades autónomas que componen la muestra española en cuanto al dominio de los atributos.

Método

Muestra

Se ha empleado una submuestra de 1.538 alumnos españoles que resolvieron el cuadernillo 1 de PISA-2006. La distribución por comunidades autónomas es: Andalucía (116), Aragón (120), Asturias (117), Cantabria (119), Castilla y León (122), Cataluña (117), Galicia (124), La Rioja (113), Navarra (124), País Vasco (302) y resto de España (164).

Procedimiento

El presente estudio se divide en cinco pasos que se explicarán enseguida. Los análisis se han llevado a cabo en el paquete *ltm* para R (Rizopoulos, 2006).

1. Obtención de la matriz de respuestas a los ítem-estímulo: partiendo de la ecuación básica del método LSDM, la respuesta correcta del ítem-estímulo requiere el dominio de todos los atributos implicados (modelo conjuntivo), por lo tanto, para obtener las respuestas de los estudiantes a cada uno de los 18 ítems-estímulo se han multiplicado las respuestas binarias de los ítems individuales correspondientes.
2. Análisis de la dimensionalidad de los ítems: se ha estudiado la dimensionalidad del conjunto de ítems-estímulo mediante un análisis factorial exploratorio, como método el de ejes principales con rotación Varimax y utilizando como entrada al análisis la matriz de correlaciones tetracóricas, indicada para ítems binarios. Los resultados indican la presencia de dos factores, por ello se ha decidido trabajar con un subconjunto de 12 ítems que conforman el primer factor.
3. Calibración y ajuste al modelo TRI de Rasch y estimación de la habilidad de los examinados.

Estímulo	PISA ID.	Formato	A ₁	A ₂	A ₃
11 Efecto invernadero	S114Q03	R.A	2	0	1
	S114Q04	R.A			
	S114Q05	R.A			
12 Ropa	S213Q01	E.M.C	0	1	1
	S213Q02	E.M			
13 Temperatura tierra	S269Q01	R.A	0	3	0
	S269Q03	R.A			
	S269Q04	E.M.C			
14 Leche	S326Q01	R.A	3	1	0
	S326Q02	R.A			
	S326Q03	E.M			
	S326Q04	E.M.C			
15 Césped avena	S408Q01	E.M	0	3	1
	S408Q03	R.A			
	S408Q04	E.M.C			
	S408Q05	E.M			
16 Paneles solares	S415Q02	E.M	0	1	2
	S415Q07	E.M.C			
	S415Q08	E.M.C			
17 La luna	S416Q01	R.E	1	0	0
18 Isla Pingüinos	S425Q02	R.A	2	1	1
	S425Q03	R.A			
	S425Q04	R.A			
	S425Q05	E.M			
19 Gran Cañón	S426Q03	E.M	0	2	1
	S426Q05	E.M			
	S426Q07	E.M.C			
110 Extinguiendo fuegos	S437Q01	E.M	0	4	0
	S437Q03	E.M			
	S437Q04	E.M			
	S437Q06	E.M			
111 Lluvia ácida	S485Q02	R.A	1	1	1
	S485Q03	E.M			
	S485Q05	R.A			
112 Ejercicio físico	S493Q01	E.M.C	0	3	0
	S493Q03	E.M.C			
	S493Q05	R.A			
113 Radioterapia	S495Q01	E.M.C	3	0	1
	S495Q02	E.M.C			
	S495Q03	R.A			
	S495Q04	R.A			
114 Cultivos modificados	S508Q02	E.M.C	0	0	2
	S508Q03	E.M			
115 Tren magnético	S510Q01	E.M.C	0	2	0
	S510Q04	R.A			
116 Desastre y desarrollo	S514Q02	R.A	2	1	0
	S514Q03	R.A			
	S514Q04	E.M.C			
117 Cocina	S521Q02	E.M	0	2	0
	S521Q06	E.M			
118 Extinción dinosaurios	S527Q01	E.M.C	1	2	0
	S527Q03	E.M.C			
	S527Q04	E.M.C			

4. Aplicación del método LSDM. Para realizar dicha aplicación se han prefijado 13 niveles de habilidad (θ) entre -3 y 3 con una distancia de 0.5. Se incluye el análisis de la recuperación de las CCI de cada ítem mediante los valores MAD y se estudian las curvas de probabilidad de los atributos.
5. Análisis de las diferencias en el dominio de los procesos entre las comunidades autónomas. Para realizar dicho análisis se ha clasificado la habilidad estimada de los sujetos en uno de los 13 puntos fijos de habilidad (el más cercano al nivel de cada alumno); de acuerdo a dicha clasificación se asigna a cada examinado la probabilidad de dominio de cada proceso, correspondiente a los resultados obtenidos en el LSDM. Posteriormente se realizan análisis de varianza (ANOVA) utilizando como variables dependientes las probabilidades de acierto de cada atributo y como factores cada una de las comunidades autónomas participantes. En los ANOVAs se ha utilizado la corrección de Bonferroni para controlar el error de tipo I.

Resultados

Análisis de la dimensionalidad, calibración y ajuste del modelo TRI de Rasch

La tabla 2 presenta las saturaciones factoriales obtenidas tras la rotación Varimax y con el método de ejes principales. Se han señalado con negrita los 12 ítems que conforman el primer factor. Los análisis posteriores se han hecho con este subconjunto de ítems unidimensionales. En la tabla 2 también se presentan los parámetros de dificultad de los ítems estimados con el modelo de Rasch, en métrica logística, y los errores estándar de cada uno de ellos. Además se incluye el estadístico χ^2 para valorar su ajuste.

No	Saturaciones		Modelo TRI Rasch				LSDM
	Factor 1	Factor 2	b	S.E	χ^2	$p(\chi^2)$	MAD
11	0.375	0.286	2.316	0.086	0.177	0.990	0.017
12	0.373	0.526					
13	0.591	0.383	2.115	0.081	1.067	0.208	0.017
14	0.721	0.218	2.732	0.097	1.227	0.366	0.014
15	0.421	0.379	2.888	0.102	0.664	0.604	0.062
16	0.257	0.696					
17	0.220	0.398					
18	0.425	0.399	3.107	0.110	0.897	0.495	0.057
19	0.462	0.355	0.133	0.062	0.493	0.950	0.244
110	0.307	0.516					
111	0.465	0.312	1.225	0.068	0.330	0.990	0.117
112	0.534	0.495	0.733	0.064	0.586	0.535	0.203
113	0.435	0.424	2.384	0.087	1.259	0.198	0.038
114	0.496	0.301	-0.289	0.062	0.581	0.901	0.047
115	0.468	0.381	1.386	0.070	0.664	0.564	0.056
116	0.367	0.477					
117	0.126	0.241					
118	0.516	0.260	2.868	0.101	1.029	0.485	0.062

Los resultados revelan que todos los ítems presentan muy buen ajuste al modelo ($p>0.05$), adicionalmente, los errores estándar son bajos indicando buena precisión en la estimación de los parámetros.

Aplicación del LSDM

La figura 1 representa las curvas de probabilidad de los atributos (CPAs) en los 13 niveles fijos de habilidad. En dicha figura se puede observar que las CPAs exhiben el comportamiento lógico esperado: a) son monótonas crecientes, indicando que a medida que aumenta la habilidad aumenta también la probabilidad de dominio de los atributos, y b) el orden de la dificultad de los atributos coincide con lo esperado de acuerdo a la teoría. El proceso más fácil ha resultado ser A₃: utilizar pruebas científicas, seguido por A₂: explicar fenómenos científicamente y el más difícil ha sido A₁: identificar cuestiones científicas.

En la columna derecha de la tabla 2 se presentan los valores MAD para los 12 ítems que se han incluido en el análisis. Se han señalado con negrita los tres ítems que están mal recuperados de acuerdo al criterio de Dimitrov (2007). El método estaría indicando que es necesario reconsiderar los vínculos entre ítems y atributos para estos tres ítems o incluso la relación conjuntiva que se ha establecido entre ellos. Las CCI de los nueve ítems restantes se encuentran bien o muy bien recuperadas, apoyando la validez de la estructura cognitiva propuesta.

Diferencias en la ejecución de los atributos entre comunidades autónomas

La tabla 3 muestra los promedios y desviaciones estándar de las probabilidades de dominio de los procesos en cada comunidad autónoma. En la tabla se puede observar que en todas las comunidades se mantiene el orden en la dificultad de los procesos ya evidenciado en la aplicación general del LSDM, el más difícil es A₃, con una proporción media de dominio de 0.26, y el más fácil es A₁, con una proporción media de 0.7.

Por medio del análisis de varianza (ANOVA) se ha analizado la magnitud de las diferencias en las probabilidades de dominio entre las comunidades. La comprobación de los supuestos de homogeneidad de varianza (prueba de Levene F= 1,92; $p>0.01$) indican

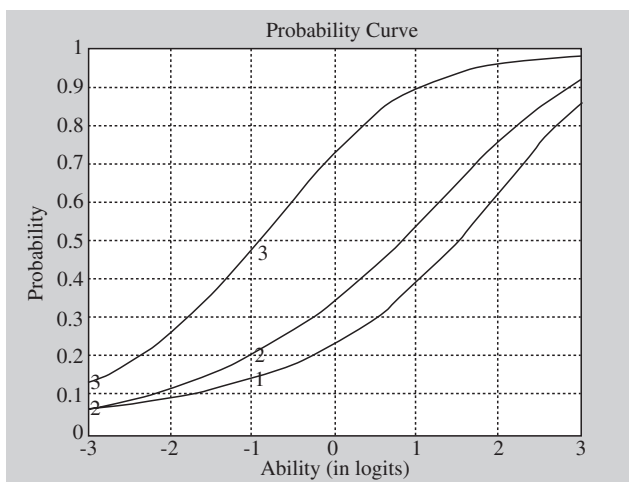


Figura 1. Curvas de probabilidad de los atributos estimados con el LSDM

que es adecuada la utilización de ANOVA en esta submuestra. Tras este análisis se ha encontrado que en el proceso *identificar cuestiones científicas* (A_1) existen diferencias significativas entre el País Vasco y las Comunidades: Aragón (diferencia= -0.05; $p<0.001$) y Navarra (diferencia= -0.06; $p<0.001$). En el atributo *usar evidencia científica* (A_2) también se han encontrado diferencias significativas entre las medias de estas tres Comunidades con una magnitud de -0.07 ($p<0.001$) entre País Vasco y Aragón y de

-0.08 ($p<0.001$) entre País Vasco y Navarra. Finalmente, se han encontrado también diferencias entre las Comunidades mencionadas en el proceso *utilizar pruebas científicas* (A_3) con una magnitud de -0.07 ($p<0.001$) en los dos contrastes.

Discusión

En el presente trabajo se ha llevado a cabo un análisis de la estructura cognitiva de 18 ítems compuestos de la escala de ciencias de PISA-2006 en la muestra española. Los resultados sugieren que la estructura de tres procesos propuesta por la OCDE para PISA-2006 explica completa y correctamente la ejecución de la mayoría de los ítems analizados. A pesar de la simplicidad de la matriz Q las CCI han sido recuperadas con precisión, excepto las curvas de los ítems 9 (Gran Cañón), 11 (Lluvia ácida) y 12 (Ejercicio físico).

En el presente estudio también se han analizado las diferencias entre las comunidades autónomas participantes en cuanto al dominio de los atributos, encontrando diferencias significativas (en los tres procesos) entre el País Vasco, Navarra y Aragón. Estos resultados merecen ser ampliados a través de la inclusión de otras competencias (por ejemplo, matemáticas y comprensión lectora).

Los resultados aquí presentados corresponden únicamente a los ítems del cuadernillo 1 aplicados a la muestra española. Por lo tanto, es importante analizar el comportamiento de la estructura cognitiva propuesta por la OCDE en todos los cuadernillos, en la muestra total de España, así como en otros países, ya que éstos permitirán, por ejemplo, tener una mejor comprensión del comportamiento de los atributos.

El método LSDM ha demostrado ser una herramienta útil y parsimoniosa para este tipo de análisis.

Comunidad autónoma	A_1		A_2		A_3	
	\bar{X}	s.e	\bar{X}	s.e	\bar{X}	s.e
Andalucía	.246	.105	.346	.136	.689	.150
Aragón	.290	.135	.400	.164	.741	.152
Asturias	.275	.124	.383	.154	.726	.152
Cantabria	.255	.117	.355	.142	.703	.138
Castilla y León	.280	.130	.387	.159	.732	.148
Cataluña	.254	.104	.357	.136	.706	.146
Galicia	.282	.144	.387	.168	.726	.151
La Rioja	.269	.112	.376	.142	.728	.143
Navarra	.298	.143	.409	.174	.741	.164
País Vasco	.236	.104	.330	.132	.672	.147
Resto de España	.259	.136	.356	.166	.684	.163
Total	.264	.124	.366	.153	.708	.152

Referencias

- DiBello, L.V., Stout, W.F., y Roussos, L.A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. En P. Nichol, S. Chipman y R. Brennan (Eds.): *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Lawrence Erlbaum.
- Dimitrov, D.M., y Raykov, T. (2003). Validation of cognitive structures: A structural equation modeling approach. *Multivariate Behavioral Research*, 38, 1-23.
- 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(5), 367-387.
- Embretson, S.E., y Reise, S. (2000). *Item Response Theory for Psychologists*. London: Lawrence Associates Publishers.
- Embretson, S.E. (1984). A general latent trait model for response processes. *Psychometrika*, 49, 175-186.
- Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- Henson, R., y Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29(4), 262-277.
- Junker, B.W., y Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272.
- MEC (2007). *PISA 2006. Programa para la Evaluación Internacional de Alumnos de la OCDE. Informe español*. Consultado el 17 de julio de 2008 desde <http://www.institutodeevaluacion.mec.es/contenidos/internacional/pisainforme2006.pdf>.
- Medina-Díaz, M. (1993). Analysis of cognitive structure using the linear logistic test model and quadratic assignment. *Applied Psychological Measurement*, 17, 117-130.
- OECD (2009). *PISA 2006: Technical Report*. París: OECD. Consultado el 17 de marzo de 2009 desde <http://www.oecd.org/dataoecd/0/47/42025182.pdf>.
- OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis*. París: OECD. Consultado el 17 de julio de 2008 desde <http://www.pisa.oecd.org/dataoecd/30/17/39703267.pdf>.
- Rizopoulos, D. (2006). Ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-23.
- Romero, S.J., Ponsoda, V., y Ximénez, C. (2006). Validación de la estructura cognitiva del test de signos mediante modelos de ecuaciones estructurales. *Psicothema*, 18, 835-840.
- Tatsuoka, K.K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, 10(1), 55-73.
- Tatsuoka, K.K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern classification approach. En P. Nichol, S. Chipman y R. Brennan (Eds.): *Cognitively diagnostic assessment* (pp. 327-360). Hillsdale, NJ: Lawrence Erlbaum.
- Von Davier, M., y Yamamoto, K. (2004). Partially observed mixtures of IRT models: An extension of the generalized partial-credit model. *Applied Psychological Measurement*, 28, 389-406.
- Whitley, S.E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45(4), 479-494.
- Whitley, S.E., y Schneider, L.M. (1981). Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement*, 5(3), 383-397.
- Yan, D., Almond, R.G., y Mislevy, R.J. (2004). *Comparisons of cognitive diagnostic models* (ETS RR-04-02). Princeton, NJ: ETS. Disponible en <http://www.ets.org/research/newpubs.html>.