

# Validación de la estructura cognitiva del test de signos mediante modelos de ecuaciones estructurales

Sonia J. Romero, Vicente Ponsoda y Carmen Ximénez  
Universidad Autónoma de Madrid

El objetivo de este trabajo es realizar un estudio de validación de los procesos cognitivos implicados en la resolución de ítems de un test de operaciones aritméticas básicas entre números enteros. La validación de la estructura cognitiva propuesta para la tarea se lleva a cabo mediante modelos de ecuaciones estructurales y triangulación. Los resultados muestran relaciones de subordinación fuertes y positivas entre los ítems en algunas de las rutas, apoyando solo parcialmente la estructura propuesta. Sin embargo, la triangulación aporta una mayor evidencia de validez.

*Validation of the cognitive structure of the test of signs by structural equation modeling.* The present work is aimed to carry out a validation study of the cognitive operations required for the correct solution of items of a math test which includes basic arithmetic operations between integer numbers. The validation of the hypothesized cognitive structure is made by means of structural equation modeling and triangulation methods. Results show strong and positive cognitive subordination relationships between some items but the structural equation model fit only provides a partial support for the proposed structure. However, the triangulation procedure provides further evidence of validity.

En las últimas décadas se ha producido un interés creciente por el desarrollo de modelos psicométricos que tienen en cuenta los Procesos Cognitivos (PCs) utilizados al resolver los ítems de un test. Un ejemplo son los denominados *modelos componenciales* (Embretson, 1999; Fischer, 1995; van der Linden y Hambleton, 1997). Estos modelos requieren: a) determinar la cantidad y orden de los PCs implicados en la resolución de los ítems; y b) proponer un modelo matemático que estime la probabilidad de acertar el ítem según sus propiedades estructurales y el nivel de habilidad del evaluado.

Uno de los modelos componenciales más utilizados es el Modelo Logístico Lineal de Rasgo Latente (LLTM) de Fischer (1973). Es una extensión del modelo de Rasch, en el que se establecen restricciones lineales en los parámetros. La descomposición lineal del parámetro de dificultad para cada ítem se expresa mediante:

$$\beta_i = \sum_{k=1}^p w_{ik} \alpha_k + c \quad (1)$$

Donde  $\alpha_k$ , con  $k= 1, \dots, p$ , son los llamados *parámetros básicos* e indican la contribución de cada PC a la explicación de las dificultades de los ítems;  $w_{ik}$  es el peso del PC  $k$  e indica si éste ha de tenerse en cuenta para la resolución del ítem  $i$ ; y  $c$  es una constante de normalización.

El LLTM fue desarrollado para analizar tests en los que se conoce la estructura de cada ítem en términos de los PCs necesarios para su correcta resolución. La estructura cognitiva se refleja en la denominada matriz de pesos (**W**) del siguiente modo: si el ítem  $i$  requiere la operación  $k$ ,  $w_{ik}= 1$ ; en caso contrario,  $w_{ik}= 0$ . Por ello, los  $w_{ik}$  dependen del diseño del ítem y se consideran fijos. Los valores de  $\alpha_k$  se estiman a partir de **W** y de las respuestas de los sujetos mediante máxima verosimilitud condicional. A continuación se estiman las dificultades mediante la ecuación (1), que serán usadas a su vez en el modelo de Rasch.

La aplicación correcta del modelo LLTM requiere tres tipos de contrastes: 1) el ajuste previo al modelo de Rasch; 2) las restricciones lineales de la ecuación (1) mediante el estadístico de la razón de verosimilitudes condicional CLR (Fischer y Ponocny-Seligler, 1998); y 3) estudios de validación del modelo cognitivo, representado formalmente en la matriz **W**.

Apenas se han propuesto procedimientos para realizar estudios de validación. Medina-Díaz (1993) propuso uno basado en técnicas de asignación cuadrática (QA), que evalúa la asociación entre la matriz estructural **C** (que describe las relaciones entre los ítems derivadas de la matriz **W**) y la matriz de proximidades **Q** (que contiene las relaciones observadas entre los ítems). El principal inconveniente del método QA es que requiere información empírica de qué operaciones se han aplicado en la solución de cada ítem, no bastando con su solución final.

Debido a estas dificultades, Dimitrov y Raykov (2003) proponen el uso de modelos de ecuaciones estructurales (SEM) para el contraste de las relaciones establecidas en la ecuación (1). El procedimiento que presentan es complementario y no alternativo, pues mientras el estadístico CLR pone a prueba si la dificultad de los ítems se descompone según la ecuación (1), el análisis SEM evalúa si las relaciones observadas entre los ítems son compatibles

con las relaciones expresadas en el modelo psicológico propuesto para explicar cómo se resuelven.

El procedimiento de validación por SEM sigue los siguientes pasos: 1) reducir la matriz **W** a la matriz **S** de subordinaciones cognitivas. Un ítem  $I_j$  está subordinado a un ítem  $I_i$  si y solo si las operaciones cognitivas requeridas para su correcta resolución son un subconjunto propio de las operaciones requeridas por el ítem  $I_i$ . En la matriz **S**,  $s_{ij} = 1$  indica subordinación del ítem  $I_j$  al  $I_i$ ; en otro caso,  $s_{ij} = 0$ . 2) Representar gráficamente las subordinaciones definidas en la matriz **S**. Cada ítem se representa con un círculo que denota la habilidad latente requerida para su correcta solución. Las flechas  $I_j$  al  $I_i$  indican subordinación del ítem  $I_j$  al  $I_i$ . 3) Evaluar el ajuste del modelo mediante los procedimientos de la metodología SEM.

Para el ítem  $i$  se asume que existe una variable latente continua ( $\eta_i$ ) que representa la habilidad requerida para su correcta solución. Existe también un valor del umbral de habilidad ( $\tau_i$ ) cuya superación implica el acierto del ítem, de forma que:

$$X_i = 1, \text{ si } \eta_i \geq \tau_i \quad \text{y} \quad X_i = 0, \text{ si } \eta_i \leq \tau_i \quad (2)$$

Partiendo de lo anterior, la ecuación estructural del modelo es:

$$n = \mathbf{B}\eta + \zeta \quad (3)$$

donde  $\eta$  es el vector de variables latentes que representan las habilidades para resolver los ítems, **B** es la matriz de relaciones entre las habilidades y  $\zeta$  el vector de residuales.

Los parámetros  $\beta$  representan el grado en que la habilidad para acertar un ítem está linealmente relacionada con la de otro(s) ítem(s) subordinados a él. Por tanto, el enfoque SEM proporciona una validación general de la estructura cognitiva y también de las relaciones de subordinación entre ítems, indicando cuáles se pueden excluir o añadir.

Además de aplicar la metodología SEM, Dimitrov y Raykov proponen la *triangulación*, un procedimiento de validación que consiste en comprobar que la dificultad de los ítems aumenta a medida que se pasa de un ítem  $I_i$  a otro  $I_k$  (si  $I_i$  está subordinado a  $I_k$ ) pues al incrementarse el número de procesos involucrados, la dificultad del ítem deberá ser mayor.

Que sepamos, no se han realizado aplicaciones posteriores de la metodología propuesta por Dimitrov y Raykov. En este trabajo se aplican ambos procedimientos, SEM y triangulación, para hacer un estudio de validación de la estructura cognitiva del test de signos (Alonso y Olea, 1997). Real, Olea, Ponsoda, Revuelta y Abad (1999) delimitaron 10 PCs para dicho test y estudiaron el ajuste al modelo de Rasch y al modelo LLTM, encontrando una elevada relación lineal (0.87) entre los parámetros de dificultad de ambos modelos, aunque el estadístico CLR indicaba desajuste del LLTM y 22 de los 66 ítems no ajustaban al modelo de Rasch. Asimismo, dicho estudio no planteaba la validación de la estructura cognitiva propuesta.

El objetivo del presente trabajo es proponer una estructura cognitiva para el test de signos y hacer un estudio de validación de la misma, aplicando el enfoque SEM. En primer lugar, se hace un estudio de validación de una estructura propuesta por Real et al. Dado que dicha estructura no es plausible, se propone una nueva estructura y se lleva a cabo su estudio de validación mediante el enfoque SEM y la triangulación.

### El test de signos: delimitación de la estructura cognitiva

El test de signos consta de 66 ítems de forma  $a*b=c$ , donde \* es una de las cuatro operaciones básicas (suma, resta, multiplicación y división) y  $a$ ,  $b$  y  $c$  son números enteros positivos y negativos de uno o dos dígitos. El test consta de 24 ítems de sumas, 24 de restas, 9 de productos y 9 de divisiones. El formato de respuesta es de «sentencia-verificación» (el niño ha de identificar la corrección o incorrección del resultado ofrecido).

En primer lugar se aplicó la metodología SEM al modelo propuesto por Real et al. Para ello, se utilizaron los datos y las especificaciones que se exponen en el apartado «Método». No fue posible llegar a una solución convergente según el criterio de Jöreskog (1967): «que la solución converja en las 250 primeras iteraciones» (p. 460). Puesto que las relaciones de subordinaciones entre ítems planteadas en dicho modelo no son plausibles se decidió plantear un nuevo modelo. Para establecer los procesos se recurrió al análisis sistemático las operaciones básicas con números enteros (Kamii, 1989; Maza, 1991, 1999), se realizaron entrevistas a profesores de matemáticas y se revisaron materiales curriculares.

Algunos autores usan el concepto «hechos numéricos» para referirse a las operaciones básicas que se enseñan en Educación Primaria (Maza, 1991; 1999). Siendo  $a$  y  $b$  números naturales, el «hecho aditivo»,  $(a+b)$ , se aprende paralelamente al «hecho sustractivo»,  $(a-b)$ , con  $a>b$ . A continuación se introduce el «hecho multiplicativo»  $(a \times b)$ , que se enseña como una extensión de la suma, pero se aplica mediante la memorización de tablas de multiplicar; y, finalmente, el «hecho divisivo» (división exacta con una sola cifra en el divisor:  $a/b=c$ , con  $a>b$ ), que consiste en convertir el cociente en el correspondiente producto ( $b \times c = a$ ).

Cuando se han afianzado las operaciones con números naturales se introducen las operaciones con números enteros. El proceso de enseñanza de la *suma* con enteros comienza con el concepto de orden, con la utilización de la recta y la noción de valor absoluto. Cuando el alumno maneja dichos conceptos se introduce la suma con números de igual signo con la instrucción: «se suman las cantidades y al resultado se le antepone el mismo signo de los números implicados», que se plasma en que el resultado de sumar dos positivos siempre es positivo, y el de dos negativos, negativo. La suma con números de diferente signo se enseña con la instrucción: «se resta el número de menor cantidad al de mayor cantidad y al resultado se le pone el signo del número de mayor cantidad». La *resta* entre números enteros se enseña utilizando el concepto de «inverso aditivo». Sea cual sea el signo de  $b$ , se debe cambiar su signo, cambiar el operador de resta por el de suma y resolver la suma. Por ejemplo:  $(-8)-(-3) = (-8)+(3) = (-5)$ . La *multiplicación* y la *división* con enteros se enseña con la instrucción: «el producto de dos números positivos es siempre positivo; el de dos negativos, positivo; y el de uno positivo con uno negativo, negativo». Derivado del análisis anterior, la tabla 1 presenta los ocho PCs que se proponen.

El siguiente paso consiste en la obtención de las matrices **W** y **S**. Dimitrov y Raykov aplicaron el SEM a un test de 15 ítems dicotómicos. En su estudio cada ítem requería un conjunto de procesos distinto al de los demás, de ahí que la estructura teórica estableciera relaciones entre ítems individuales. En nuestro caso, puesto que hay ítems que comparten los mismos PCs, se decidió aplicar el denominado «ítem parceling» (Kishton y Widaman, 1994), que consiste en formar grupos de ítems tomando como puntuación la suma de las puntuaciones en sus ítems, y aplicar el análisis SEM a dichos grupos en lugar de a los ítems individuales. Es-

te procedimiento sólo se justifica cuando los grupos tienen sentido teórico y sus ítems son unidimensionales. Según esto, hemos obtenido que los 66 ítems dan lugar a 11 grupos de ítems unidimensionales (véase última columna de la tabla 2) que tienen sentido teórico puesto que requieren los mismos procesos.

La parte superior de la tabla 2 contiene la matriz **W**, que muestra los procesos requeridos por cada grupo de ítems. Como ilustración, el ítem 55,  $(-8)+(-5) = -3$ , que pertenece al grupo G2, cuya respuesta ofrecida es incorrecta, requiere aplicar el PC1 (suma) y el PC7 (signo negativo). El ítem 52,  $(6)+(2) = 8$ , que pertenece al

grupo G1, requiere aplicar sólo el PC1. Por tanto, sus PCs son un subconjunto propio de los del ítem 55 y está subordinado a él. La matriz **S** muestra las subordinaciones entre los grupos de ítems (véase parte inferior de la tabla 2). Por ejemplo,  $s_{21} = 1$  indica que los ítems del grupo G1 están subordinados a los del grupo G2.

Las relaciones de subordinación entre grupos de ítems definidas en la matriz **S** aparecen expresadas gráficamente en la parte superior de la figura 1. Como puede verse, hay dos casos independientes de relaciones entre grupos de ítems de suma y resta: el caso 1, con subordinaciones entre cuatro grupos de ítems; y el ca-

Operación	Descripción del proceso
PC1 (Suma)	Suma entre dos números naturales
PC2 (Resta)	Resta entre dos números naturales, con $a > b$
PC3 (Producto)	Multiplicación entre dos números naturales
PC4 (Cociente)	División exacta entre dos números naturales
PC5 (Identificación de mayor y menor en valor absoluto)	Identificación del componente mayor en valor absoluto y planteamiento de la resta del menor al mayor
PC6 (Intercambio de componentes)	Intercambio de posición entre $a$ y $b$
PC7 (Signo del resultado)	Determinación del signo positivo o negativo del resultado aplicando la correspondiente regla
PC8 (Conversión de resta en suma)	Cambio del signo de $b$ y del operador de resta por suma

Características	Matriz W								T
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	
G1 $(a)+(b)$	1	0	0	0	0	0	0	0	0.86
G2 $(-a)+(-b)$	1	0	0	0	0	0	1	0	0.95
G3 $(-a)+(b), (a)+(-b) a > b$	0	1	0	0	1	0	1	0	0.99
G4 $(-a)+(b), (a)+(-b) a < b$	0	1	0	0	1	1	1	0	0.96
G5 $(a)-(b) a > b$	0	1	0	0	0	0	0	0	0.99
G6 $(-a)-(-b) a > b$	0	1	0	0	1	0	1	1	0.99
G7 $(-a)-(-b), (a)-(-b) a < b$	0	1	0	0	1	1	1	1	0.94
G8 $(-a)-(-b)$	1	0	0	0	0	0	1	1	0.99
G9 $(a)-(-b)$	1	0	0	0	0	0	0	1	0.99
G10 $(-a) \times (b), (a) \times (-b), (-a) \times (-b)$	0	0	1	0	0	0	1	0	0.95
G11 $(-a)/(b), (a)/(-b), (-a)/(-b)$	0	0	0	1	0	0	1	0	0.98

Matriz S											
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	0	0	0	0	0	0	0	0	0	0	0
G2	1	0	0	0	0	0	0	0	0	0	0
G3	0	0	0	0	1	0	0	0	0	0	0
G4	0	0	1	0	1	0	0	0	0	0	0
G5	0	0	0	0	0	0	0	0	0	0	0
G6	0	0	1	0	1	0	0	0	0	0	0
G7	0	0	1	1	1	1	0	0	0	0	0
G8	1	1	0	0	0	0	0	0	1	0	0
G9	1	0	0	0	0	0	0	0	0	0	0
G10	0	0	0	0	0	0	0	0	0	0	0
G11	0	0	0	0	0	0	0	0	0	0	0

Donde T es el índice de unidimensionalidad de Tanaka ofrecido por NOHARM (Fraser, 1988)

so 2 con cinco. Los grupos G10 y G11 no aparecen representados porque no tienen relaciones de subordinación con el resto de los ítems, pues el test de signos no tiene multiplicaciones y divisiones con números naturales.

Validación de la estructura cognitiva propuesta mediante SEM

Método

Instrumento y participantes

El test de signos fue aplicado a una muestra de 221 alumnos de 7º de EGB y 1º de la ESO de tres colegios de Madrid.

Procedimiento

En primer lugar se realizó el análisis SEM para cada modelo path planteado en la figura 1. El método de estimación empleado fue WLS, que ajusta el modelo a la matriz de correlaciones policóricas y requiere el cálculo de la matriz de covarianzas asintóticas. Para la evaluación del ajuste del modelo, además del estadístico chi-cuadrado de Pearson y su prueba de significación, se emplea el criterio mixto de Hu y Bentler (1999): se considera buen ajuste cuando el índice RMSEA <0.05 y el índice CFI >0.95. También se examinaron los residuos estandarizados. Los análisis SEM se han realizado con los programas PRELIS 2 y LISREL 8.71. Para propósitos de triangulación, se compara la proporción media de aciertos en cada grupo de ítems.

Resultados

La tabla 3 incluye los estadísticos descriptivos, la matriz de correlaciones policóricas y la matriz de residuos estandarizados.

*Ajuste de los modelos SEM:* la tabla 4 presenta los índices de ajuste, los parámetros estimados en el modelo y sus respectivos errores típicos (entre paréntesis). Como puede verse, ambos modelos ofrecen índices de ajuste aceptables. En el caso 1, para obtener un buen ajuste, ha sido necesario reespecificar el modelo añadiendo la correlación entre los errores de los grupos G2 y G9 (véase panel inferior izquierdo de la figura 1). Asimismo, sólo hay dos residuos significativos: en el caso 1 el modelo predice mal la covarianza entre los grupos G1 y G2 y en el caso 2 entre G4 y G5 (véase tabla 3). Dado que ambos modelos obtienen un ajuste aceptable, es adecuado proceder a su interpretación.

En el caso 1 se aprecia una relación fuerte de subordinación G9→G8, posiblemente debida a que ambos grupos requieren los mismos procesos excepto el PC7. La relación G2→G8 es también positiva y significativa, indicando que para resolver los ítems del grupo G8 es importante la habilidad empleada en los ítems del grupo G2. No han resultado significativas las relaciones G1→G2 y G1→G9. El grupo G1 está formado exclusivamente por ítems del tipo (a)+(b), que han resultado muy fáciles (la proporción media de acierto es 0.98). No es de extrañar entonces su escasa capacidad predictora. Existe una relación moderada (0.23) que no se esperaba entre los residuos asociados al rendimiento en los grupos G2 y G9. Por tanto, otro u otros componentes distintos del requerido por G1 (que sólo requiere PC1: suma) parecen afectar al rendimiento en los ítems de ambos grupos. Una posible explicación a este resultado puede deberse a características comunes de formato pues hemos comprobado que, en todos los ítems de G2 y G9, el resultado que proporciona el ítem es incorrecto.

En el caso 2 se aprecia una relación lineal positiva y significativa en las relaciones G6→G7, G3→G4 y G4→G7. Cabía esperar dichos resultados porque los procesos requeridos por los ítems de G3 representan la mayor parte (todos excepto el PC6) de los requeridos por los de G4, y lo mismo ocurre entre los ítems de G6 y

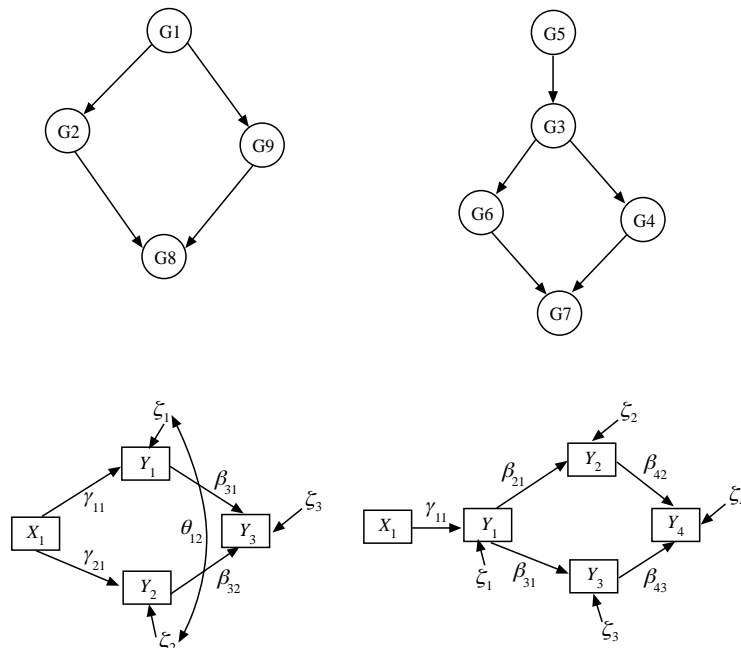


Figura 1. Diagramas path de subordinaciones entre grupos de ítems. Los gráficos de la parte superior representan los modelos de estructura cognitiva planteados para el caso 1 (a la izquierda) y 2 (a la derecha); y los de la parte inferior el diagrama path correspondiente al modelo que se ha ajustado (donde: X<sub>1</sub>= G1, Y<sub>1</sub>= G2, Y<sub>2</sub>= G9 e Y<sub>3</sub>= G8 para el caso 1; y X<sub>1</sub>= G5, Y<sub>1</sub>= G3, Y<sub>2</sub>= G6, Y<sub>3</sub>= G4 e Y<sub>4</sub>= G7 para el caso 2)

	Caso 1				Caso 2				
	G1	G2	G8	G9	G3	G4	G5	G6	G7
G1		0.49	-0.18	-0.19	G3		0.13	-0.18	0.09
G2	6.32		0.42	0.19	G4	0.92	0.26	0.01	0.20
G8	-1.10	-0.19		0.56	G5	1.24	2.64	0.13	-0.06
G9	0.35	0.07	-0.12		G6	0.34	1.42	1.65	0.74
					G7	0.22	0.07	-0.82	-0.21
Nº ítems	6	6	6	6	5	7	3	3	6
M.	5.90	4.76	3.38	2.63	4.67	6.43	2.76	1.79	3.55
D.T.	0.51	1.36	1.67	2.53	0.83	1.15	0.59	1.24	1.97
$\bar{p}$	0.98	0.79	0.56	0.44	0.93	0.92	0.93	0.60	0.59
( $S_p$ )	(0.01)	(0.08)	(0.22)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)	(0.03)

Los elementos por encima de la diagonal de la matriz son las correlaciones policóricas y por debajo los residuos estandarizados. En las filas inferiores aparecen la media (M.), desviación típica (D.T.) y proporción media de aciertos ( $\bar{p}$ ) para cada grupo de ítems

G7. Los ítems de G4 y G7 comparten también la mayoría de procesos, excepto el PC8. La relación G5→G3 no resultó significativa, a lo que puede haber contribuido el hecho de que ambos comparten un proceso y difieren en dos. Por su parte, los grupos G3 y G6 difieren únicamente en el PC8 y existe entre ellos una relación significativa pero negativa. Es decir, los niños con alto rendimiento en los ítems  $(-a)+(b)$  o  $(a)+(-b)$ , tienen un rendimiento bajo en los ítems  $(-a)-(-b)$ . Estos resultados evidencian la importancia de la operación de conversión de la resta en suma (PC8) en la resolución de este tipo de tarea y puede servir a los docentes para hacer énfasis en dicho proceso durante la enseñanza del tema.

**Triangulación:** al realizar una triangulación lógica con las proporciones medias de acierto se observa un incremento de la dificultad a medida que los grupos de ítems van añadiendo procesos. En el caso 1 se pueden hacer cinco comparaciones (G1-G2, G1-G8, G1-G9, G2-G8 y G9-G8). La tabla 3 contiene las proporciones medias de acierto de cada grupo de ítems. Puede comprobarse que las diferencias entre las proporciones medias son 0.19, 0.42, 0.54, 0.23 y -0.12, respectivamente. En cuatro de las cinco comparaciones, el aumento de procesos lleva aparejado un decremento importante en la proporción media de acierto. En el caso 2, las posibles comparaciones son nueve (G5-G3, G5-G6, G5-G7, G5-G4, G3-G6, G3-G7, G3-G4, G6-G7 y G4-G7) y las diferencias obtenidas son 0.0, 0.33, 0.34, 0.01, 0.33, 0.34, 0.01, 0.01 y 0.33, respectivamente. Se aprecia que cuatro de las nueve comparaciones no producen una diferencia reseñable. Las cinco restantes sí muestran decrementos en la proporción de acierto en los grupos de ítems con más procesos.

Si se observa la proporción de aciertos de los ítems de los distintos grupos, se aprecia que la desviación típica en los ítems del grupo G8 es casi tres veces superior a la del siguiente grupo de más variabilidad (G2). De los seis ítems del grupo G8, tres resultaron difíciles (proporciones de acierto: 0.36, 0.35 y 0.38) y tres fáciles (0.80, 0.78 y 0.71). Se observa que en los ítems difíciles ocurre que  $a>b$ , mientras que en los fáciles que  $a<b$ , lo que sugiere que el tamaño relativo de  $a$  y  $b$  puede tener un papel en este grupo de ítems, pese a no haberse tenido en cuenta en el modelo. Este resultado indica que hay factores no tenidos en cuenta por el modelo que pueden estar afectando a los resultados.

Caso 1		Caso 2	
$\gamma_{11}$ (G1→G2)	0.08 (0.43)	$\gamma_{11}$ (G5→G3)	0.00 (0.11)
$\gamma_{21}$ (G1→G9)	-0.12 (0.69)	$\beta_{21}$ (G3→G6)	-0.15 (0.07)*
$\beta_{31}$ (G2→G8)	0.30 (0.07)*	$\beta_{31}$ (G3→G4)	0.65 (0.13) *
$\beta_{32}$ (G9→G8)	0.55 (0.03) *	$\beta_{42}$ (G6→G7)	0.71 (0.07) *
		$\beta_{43}$ (G4→G7)	0.22 (0.09) *
$\zeta_1$	1.69 (0.15)*	$\zeta_1$	0.59 (0.17)*
$\zeta_2$	6.42 (0.28)*	$\zeta_2$	1.51 (0.07)*
$\zeta_3$	1.52 (0.13)*	$\zeta_3$	0.61 (0.12)*
$\theta_{12}$	0.23 (0.20)*	$\zeta_4$	1.91 (0.17)*
$\chi^2$ (p-valor)	3.07 (p= 0.08)		13.02 (p= 0.02)
gl	1		5
RMSEA	0.097		0.085
(intervalo al 90%)	(0.00-0.22)		(0.03-0.14)
CFI	0.99		0.97

Los parámetros de las variables observadas aparecen estandarizados y los de las latentes, en directas  
\*  $p<0.01$

## Discusión y conclusiones

Este estudio aporta algunas evidencias sobre la validez del modelo propuesto. Desde la idea de validez que propuso Messick (1989) un test «nunca está validado» al ser siempre posible realizar estudios complementarios sobre el significado de sus puntuaciones (en el área de aprendizaje sirva como ejemplo el estudio de la validez de la escala SIACEPA de Barca, Peralbo y Breñilla, 2004; y los estudios de validez de la evaluación del potencial de aprendizaje de Calero, 2004). En este sentido, el presente estudio proporciona información adicional sobre lo que mide el test de signos y evidencias diferentes a las encontradas tradicionalmente en estudios de validación, pues se centra en la adecuación de los PCs propuestos.

El modelo propuesto es más adecuado que el descrito en Real et al., pues se ha conseguido ajustar el modelo SEM, mientras no fue posible hacerlo al modelo propuesto por dichos autores. Sin embargo, el ajuste ha precisado de una reespecificación, ausente en el modelo de partida, y es sólo parcialmente satisfactorio. Además, los ítems de las operaciones de multiplicación y división no han sido incluidos en el modelo, al no haberse podido establecer relaciones de subordinación que les afecten. Una segunda estrategia de validación, la triangulación, ofrece resultados más acordes con lo esperado, pero no en todas las comparaciones el incremento de PCs se traduce en un incremento de la dificultad.

En el trabajo de Real et al., el modelo propuesto surgió en parte de un análisis lógico de los posibles pasos necesarios para la resolución de los ítems. El presente trabajo ha dado un paso adelante al proponer un modelo basado en cómo explican los docentes la realización de estas operaciones a los alumnos. Presumiblemente, la explicación que da el profesor sobre cómo hacer las operaciones debe estar más cerca de la posible actuación del alumno que un análisis lógico de la misma, por plausible que resulte. Por otro lado, parece poco probable que un niño, por ejemplo, ante el ítem  $(-2)-(-4)=-6$  aplique la secuencia de cinco PCs que supone el modelo, cuando su tarea es decir si el resultado es correcto o no, sin tener que dejar constancia del desarrollo de la operación. Por tanto, el siguiente paso debiera ser elaborar un modelo que se inspire en cómo resuelven esta tarea los niños, utilizando para ello ítems de respuesta construida y/o entrevistas. También podrían emplearse ítems de opción múltiple.

Futuros estudios deben continuar investigando la adecuación del enfoque SEM con respecto a otras técnicas de validación. La principal ventaja de este procedimiento es que no requiere infor-

mación empírica sobre si el niño ha aplicado o no cada proceso en el ítem. Basta con saber si ha resuelto bien o mal el ítem, lo que constituye un ahorro en tiempo y costes. Otra ventaja del enfoque SEM es que no está ligado a los supuestos específicos del LLTM o de los modelos de la teoría de respuesta al ítem. No obstante, requiere el cumplimiento de los supuestos del SEM. En cuanto a las limitaciones, la principal es que la matriz de subordinaciones  $S$  no tiene en cuenta cuántos y qué procesos son comunes a los ítems (un ítem está subordinado a otro tanto si comparte un sólo PC como si comparte varios). Es decir, el enfoque propuesto intenta validar un aspecto de la relación entre los ítems –la subordinación, tal como se ha definido–, quedando otras posibles relaciones fuera de consideración. Por ejemplo, dos modelos que utilicen los mismos procesos para la resolución de los ítems, pero aplicados en distinto orden, son indistinguibles y proporcionan los mismos resultados, pues ambos generan la misma matriz  $W$ .

Otra línea de investigación futura podría ser llevar a cabo estudios de simulación sobre la adecuación del enfoque SEM para validar estructuras cognitivas. Por ejemplo, se puede controlar el grado de inespecificación de la matriz  $W$  y apreciar cómo repercute en el ajuste del modelo y en la significación de los parámetros  $\beta$ . Por último, se necesita más investigación para analizar las ventajas e inconvenientes que introduce el «ítem parceling» en el procedimiento de validación por SEM.

#### Agradecimientos

Queremos agradecer a Julio Olea y Jesús Alonso por proporcionar los datos. Este trabajo está financiado por el proyecto de la DGICYT (ref. SEJ2004-05872).

#### Referencias

- Alonso, J., y Olea, J. (1997). Modelos de evaluación de los conocimientos matemáticos. En J. Alonso (ed.): *Evaluación del conocimiento y su adquisición*. Madrid: CIDE.
- Barca, A., Peralbo, M., y Brenlla, J.C. (2004). Atribuciones causales y enfoques de aprendizaje: la escala SIACEPA. *Psicothema*, 16, 94-103.
- Calero, M.D. (2004). Validez de la evaluación del potencial de aprendizaje. *Psicothema*, 16, 217-221.
- Dimitrov, D.M., y Raykov, T. (2003). Validation of cognitive structures: A structural equation modeling approach. *Multivariate Behavioral Research*, 38, 1-23.
- Embretson, S. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, 64, 407-433.
- Fischer, G.H. (1973). The Linear Logistic Test Model as an instrument in educational research. *Acta psychologica*, 37, 359-374.
- Fischer, G.H. (1995). The Linear Logistic Test Model. En G.H. Fischer e I.W. Molenaar (eds.): *Rasch Models: Foundations, recent developments and applications*. New York: Springer-Verlag.
- Fischer, G.H., y Ponocny-Seliger, E. (1998). *Structural Rasch modeling, handbook of the usage of LPCM-Win*. Holanda: ProGAMMA.
- Fraser, C. (1988). *NOHARM* [Computer software and manual]. Armidale, New South Wales, Australia: author.
- Hu, L., y Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32, 443-477.
- Kamii, C. (1989). *Reinventando la aritmética II*. Madrid: Visor.
- Kishton, J., y Widaman, K. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, 54, 757-765.
- Maza, C. (1991). *Enseñanza de la multiplicación y la división*. Madrid: Síntesis.
- Maza, C. (1999). *Enseñanza de la suma y de la resta*. Madrid: Síntesis.
- Medina-Díaz, M. (1993). Analysis of cognitive structure using the linear logistic test model and quadratic assignment. *Applied Psychological Measurement*, 17, 117-130.
- Messick, S. (1989). Validity. En R.L. Linn (ed.): *Educational measurement* (3ª ed.). New York: American Council on Education and Macmillan.
- Real, E., Olea, J., Ponsoda, V., Revuelta, J., y Abad, F.J. (1999). Análisis de la dificultad de un test de matemáticas mediante un modelo composicional. *Psicológica*, 20, 121-134.
- Van der Linden, W., y Hambleton, R. (1997). *Handbook of modern item response theory*. New York: Springer.