

Formulación y validación de un modelo logístico lineal para la tarea de adición y sustracción de fracciones y números mixtos

Alicia López Jáuregui y Paula Elosua Oliden
Universidad del País Vasco

En este trabajo se analizan los componentes cognitivos implicados en la realización de una tarea matemática de rendimiento, concretamente la adición y sustracción de fracciones y números mixtos, mediante el modelo logístico lineal de Fisher (LLTM, 1973). La validación de la estructura cognitiva propuesta, reflejada en la matriz de pesos de los componentes, se lleva a cabo mediante el análisis QA (asignación cuadrática). Los resultados confirman que dicha matriz describe adecuadamente los componentes cognitivos requeridos para resolver los ítems del test.

Formulation and validation of a linear logistic model for a task of fraction and mixed numbers addition and subtraction. The aim of this study is to analyze the cognitive components involved in a mathematical performance task, namely the addition and subtraction of fractional and mixed numbers, using Fisher's linear logistic trait model (LLTM). The validation of the cognitive structure proposed, reflected in the component weight matrix, is performed with the quadratic assignment analysis (QA). The results confirm that the weight matrix describes adequately the cognitive components required to solve the test items.

En los últimos años se ha producido una aproximación entre la Psicometría y la investigación cognitiva. La construcción de tests y el diseño de ítems se han visto influenciados por los nuevos planteamientos cognitivos, dando lugar a lo que se ha dado en llamar «nueva generación de tests» (Frederiksen, Mislevy, y Béjar, 1993). El mismo concepto de validez se ha visto modificado y enriquecido. La noción tradicional de la validez basada en las relaciones del test con otras medidas se complementará con el concepto de validez como propiedad fundamentada en la determinación de los procesos, estrategias y estructuras de conocimiento implicados en las respuestas a los ítems o «representación del constructo» (Embretson, 1983).

Desde esta perspectiva de integración de las dos disciplinas se han venido desarrollando una gran cantidad de modelos derivados de la Teoría de Respuesta al Ítem, que incorporan variables de contenido cognitivo en su formulación, permitiendo así la cuantificación del impacto de las mismas en las propiedades psicométricas de los tests. Todos ellos permiten transcender el mero escalamiento de sujetos e ítems ahondando algo más en la naturaleza del constructo medido. Para emplear el término acuñado por Prieto (Prieto y Delgado, 2000) son modelos *centrados en la representación*.

Sin ánimo de exhaustividad, con el propósito de ilustrar la gran profusión de modelos formulados, distinguiremos cuatro clases de modelos derivados de la Teoría de Respuesta al Ítem, relevantes

para la validez, sobre la base de la naturaleza de las aplicaciones a las que principalmente se han asociado a saber, complejidad de los factores y diseño de los ítems, medición de procesos, identificación de clases latentes y medición del cambio.

Una primera categoría estaría integrada por aquellos modelos básicamente orientados a determinar las fuentes de dificultad de los ítems recogiendo el impacto de las características de los mismos sobre el procesamiento y sobre las respuestas. El modelo logístico lineal de rasgo latente LLTM (Fisher, 1973; 1977) y el modelo logístico lineal de rasgo latente con dos parámetros LLTM 2PL (Embretson, 1999) pertenecen a este grupo. Un segundo grupo lo constituyen los modelos encaminados a estudiar la descomposición de procesos cognitivos, bien sea utilizando sub tareas para medir los componentes como es el caso del modelo multicomponencial de rasgo latente, MLTM (Whitely, 1980), el modelo general de rasgo latente, GLTM (Embretson, 1984), el modelo MIRID (Butter y De Boeck, 1998), bien sea incluyendo procedimientos de estimación (Mañá, 1995) que permiten prescindir de sub tareas, como el modelo general de rasgo latente para procesos encubiertos GLTM-CP (Embretson, 1995b). El tercer grupo incluye los modelos que permiten la identificación de clases o grupos de sujetos que difieren, bien en las estrategias de respuesta, bien en los estados de conocimiento; el modelo MIRA (Rost, 1990), el «Rule Space» (Tatsuoka, 1983), Hybrid (Yamamoto y Gitomer, 1993) y el formulado por Mislevy y Verhelst (1990) se encuentran en esta categoría. Por último se han desarrollado una serie de modelos adecuados para abordar el tratamiento del cambio. En este grupo se encuadran dos generalizaciones del LLTM, el modelo logístico lineal con supuestos relajados, LLRA (Fisher, 1977; Fisher, 1983) y el LLRA Híbrido (Fisher, 1977), el modelo multidimensional para aprendizaje y cambio, MRMLC (Embretson, 1991) y la extensión de este último, el MRMLC+ (Embretson, 1995a).

Fecha recepción: 28-11-01 • Fecha aceptación: 27-5-02

Correspondencia: Alicia López Jáuregui

Facultad de Psicología

Universidad del País Vasco

20018 San Sebastián (Spain)

E-mail: psplojaa@sc.ehu.es

Ciñéndonos a los modelos que permiten establecer un vínculo entre las características del diseño de los ítems, sus características psicométricas y los atributos cognitivos objeto de la medición, uno de los primeros en formularse y que ha sido utilizado en un gran número de aplicaciones (Spada y May, 1982; Embretson, 1985; Spada y McGaw, 1985; Hornke y Habon, 1986; Embretson y Wetzel; 1987; Smith, Kramer y Kubiak, 1992; Medina Díaz, 1993; Real, Olea Ponsoda, Revuelta y Abad, 1999) es el modelo logístico lineal de rasgo latente, LLTM (Fisher; 1973, 1977). El LLTM posibilitará el análisis del tipo de conocimientos o procesos requeridos para resolver los ítems de un test.

En este trabajo se analizarán los componentes cognitivos implicados en la realización de una tarea de adición y sustracción de fracciones y números mixtos mediante el LLTM. La importante cuestión acerca de la validez de la estructura cognitiva propuesta, reflejada en la matriz de pesos de los componentes, que en este modelo se formula apriorísticamente, se abordará mediante el análisis de asignación cuadrática (QA) (Hubert y Schultz, 1976; Hubert, Golledge, Costanzo, Gale y Halperin, 1984).

LLTM

Desde el punto de vista formal el LLTM es una extensión del modelo de Rasch (1960) o modelo logístico de un parámetro (RM), en el que se imponen ciertas restricciones lineales a los parámetros de los ítems. Estas restricciones permitirán la especificación de las contribuciones de cada operación a la dificultad del ítem. El modelo LLTM asume que los parámetros de dificultad β_i del modelo de Rasch pueden descomponerse en sumas ponderadas de los parámetros básicos a_k , formalmente:

$$\beta_i = \sum_{k=1}^p w_{ik} a_k + c \tag{1}$$

Donde,

β_i es el parámetro de dificultad del *i*-ésimo ítem en el modelo de Rasch

a_k , $k=1, \dots, p$, son los parámetros básicos del LLTM y corresponden a las dificultades de cada componente *k*

w_{ik} son los pesos dados de los parámetros básicos a_k , y representan la complejidad correspondiente al ítem *i* en el componente *k*-ésimo

c es una constante de normalización

Para cada ítem *i* el investigador define a priori un vector de pesos $w_i = (w_{i1}, w_{i2}, \dots, w_{ik}, \dots, w_{im})$ donde el elemento *k* se refiere a la ocurrencia de la operación *k* en la respuesta. La matriz formada por todos los vectores de los ítems componen la matriz **W** habitualmente compuesta de unos y ceros, en función de la presencia o ausencia de una operación determinada en la resolución de un ítem. A partir de **W** y de las respuestas de los sujetos, se estimarán los valores de a_k que servirán para calcular las dificultades β_i mediante la fórmula (1).

La aplicación del LLTM requerirá de un modelo cognitivo en el marco del cual se haga posible explicar los requerimientos de procesamiento de los ítems responsables de la variabilidad de la dificultad entre los mismos. La adecuación del LLTM se fundamenta en la plausibilidad del modelo cognitivo, representado formalmen-

te en la matriz **W** de pesos, indicadores de la carga de dificultad de cada componente en el ítem. La validación de **W** es un requerimiento previo sin el cual la estimación de los parámetros no tiene valor ya que las estimaciones contendrán información válida sólo en el caso de que **W** refleje fielmente el modo en que los sujetos resuelven los ítems (Fisher, 1995). En definitiva, la adecuación del LLTM depende estrechamente de la adecuación de **W** (ítems x componentes).

Sin embargo, un aspecto problemático señalado por Fisher (1995) es la escasez de procedimientos de que disponemos para evaluar la matriz de pesos **W**. La utilización de la mera comprobación del ajuste global del modelo mediante el test estadístico habitualmente utilizado para tal fin, el test de la razón de verosimilitud constituye un enfoque limitado en la evaluación de su validez. Es preciso la utilización de otros procedimientos complementarios encaminados a validar la matriz estructural de los componentes.

En nuestro trabajo este problema se abordará mediante la aplicación de las técnicas de asignación cuadrática (QA) (Hubert y Schultz, 1976; Hubert, Golledge, Costanzo, Gale y Halperin, 1984).

Asignación cuadrática (QA)

El análisis QA consiste en un conjunto de estrategias inicialmente aplicadas al ámbito del análisis espacial aunque posteriormente se han extendido a campos tales como la educación y la sociología. Medina Díaz (1993) mostró su utilidad en el campo de la validación de pruebas, aunque no se han realizado posteriores aplicaciones en esta dirección, a pesar de haber mostrado a nuestro juicio su utilidad. Mediante el análisis QA se evalúa la asociación entre dos matrices que contienen algún tipo de medida empírica de proximidad (correlaciones, distancias, coeficientes de emparejamiento, etc.) referidas a los mismos pares de elementos. Por una lado la matriz «estructural» (C) basada en la hipótesis de investigador acerca de la organización subyacente a los datos, y por otro la matriz empírica (Q) que refleja los patrones de respuesta efectivamente observados en la muestra.

La similitud estructural entre los mismos se operativiza mediante el cómputo del número de componentes que comparten (Medina Díaz, 1993).

Siendo *n* el número de ítems del test, la matriz teórica $C_{n \times n}$ estará formada por elementos c_{ij} que representarán la «proximidad» entre los ítems *i* y *j* y se calcularán en base a a la matriz de pesos **W** mediante la expresión,

$$c_{ij} = \sum_{k=1}^m (w_{ki} w_{kj}) \tag{2}$$

Donde,

w_{ki} y w_{kj} : pesos de los *k*-ésimos componentes para los ítems *i* y

j

m: número de componentes

A su vez la matriz $Q_{n \times n}$ contiene información empírica basada en las respuestas de los examinados a cada par de ítems. Su creación exige el análisis minucioso de cada una de las respuestas dadas por los sujetos para determinar los componentes o reglas de producción efectivamente empleadas por cada sujeto y para cada ítem. Cada elemento de Q, q_{ij} , se calculará mediante,

$$q_{ij} = \sum_{j=1}^N \sum_{k=1}^m (d_{ki} d_{kj}) \quad (3)$$

Donde,

d_{ki} y d_{kj} son los valores de las variables dicotómicas (1,0) asociadas a cada componente

m : número de componentes

N : número de sujetos

Como índice de asociación entre las dos matrices se utilizará el índice «gamma» de Hubert (Hubert, 1985) que se establece a partir de los productos cruzados de los elementos de Q y C

$$\Gamma_{QC} = \sum_i \sum_j q_{ij} c_{ij} \quad (4)$$

Método

En primer lugar trataremos de delimitar los componentes o procesos elementales definitorios de la competencia en el dominio de las operaciones con fracciones. Una vez propuesto el modelo cognitivo, será posible su integración con el modelo de medida con la consecuente posibilidad de vincular tanto el logro como las diferencias en la dificultad de los ítems a explicaciones sustantivas. Es decir, se tratará de vincular el contenido del ítem a los atributos psicológicos que pretende medir. En función de los componentes se abordará la construcción de una prueba que mida las habilidades requeridas. En este sentido seguimos las pautas establecidas por Embretson (1994) en relación con el procedimiento requerido para el desarrollo de los tests.

El proceso de resolución de la tarea de adición y sustracción de fracciones y números mixtos puede representarse como el resultado de la aplicación de una determinada serie de reglas condicionales, siendo cada una de ellas un componente de la habilidad total. Esta conceptualización que postulamos para la definición y especificación de los componentes procede de la investigación cognitiva en resolución de problemas, derivada de los modelos computacionales de procesamiento y aprendizaje (Anderson, 1976; Newell y Simon, 1972; Simon, 1978). En base a un análisis previo de la tarea se propondrán una serie de reglas u operaciones asumiéndose que cada una de ellas modelará una parte del conocimiento necesario para la resolución de los ítems.

Con el fin de construir los ítems del test y teniendo en cuenta que para la construcción de una prueba de rendimiento no sólo es relevante la habilidad a evaluar, sino la población potencialmente destinataria de la misma y su nivel instruccional, recurrimos al análisis sistemático del proceso de resolución de la adición y sustracción de fracciones y números mixtos (Tatsuoka, Birenbaum, Yamamoto, y Dowd, 1984), así como a la información procedente de las entrevistas a profesores de matemáticas de E.S.O. y a la revisión de diferentes materiales curriculares (Vicens Vives, 1997; Erein, 1995).

A partir de este análisis se establecieron siete reglas de producción, cada una de las cuales contendrá parte del conocimiento necesario para resolver este tipo de problemas. La especificación de cada una de ellas aparece en la tabla 1.

En razón de las reglas de producción establecidas se construyó un test compuesto por 21 ítems de respuesta dicotómica. Los ítems incluidos en la prueba demandan diferentes combinaciones de los componentes cognitivos modelizados por las siete reglas de producción establecidas. En la tabla siguiente se presenta el número de ítem, su enunciado y una secuencia de unos y ceros, que representan el requerimiento (1) o no requerimiento (0) de cada una de las reglas de producción.

Muestra

Según la actual programación de contenidos didácticos a lo largo del ciclo educativo la instrucción en las operaciones con fracciones y números mixtos tiene lugar entre los cursos de primero, segundo y tercero de Educación Secundaria Obligatoria (E.S.O.) por lo que la muestra fue seleccionada entre alumnos de dichos cursos. La componen 771 sujetos con edades comprendidas entre los 10 y 12 años pertenecientes a cinco centros educativos de la Comunidad Autónoma Vasca que accedieron a colaborar con el estudio. El 55,5% son niños y el 44,5% niñas.

Además de pedir a los sujetos que cumplimentaran las hojas de respuesta estandarizadas se les solicitó que desarrollaran su procedimiento de solución completo en hojas aparte. La elección de esta estrategia se debió al interés en obtener información cualitativa pormenorizada acerca de la resolución de las tareas propuestas.

Resultados

Descriptivos y fiabilidad

La tabla 3 muestra las medias aritméticas y varianzas por cursos y la media y varianza total.

Tabla 1
Reglas de producción

- | |
|---|
| <ol style="list-style-type: none"> 1. SI las dos fracciones b/c, e/f a sumar (restar) tienen denominadores diferentes y primos entre sí, ENTONCES se multiplican los denominadores para obtener el denominador del resultado y se multiplica cada numerador por el denominador de la otra fracción bf / cf ± ce / cf 2. SI las dos fracciones b/c, e/f a sumar (restar) tienen los denominadores diferentes y uno de ellos (c) es múltiplo del otro (f) ENTONCES "c" se establece como común denominador, se divide c entre f y el resultado se multiplica por el numerador del segundo dejando invariable el numerador del primero. 3. SI las dos fracciones b/c, e/f a sumar (restar) tienen los denominadores diferentes no primos entre sí y con mínimo común múltiplo no coincidente con ninguno de ellos, ENTONCES se establece el común denominador a partir del mínimo común múltiplo (m), se efectúa m/c×b y m/f×e. 4. SI uno de los componentes de la suma (resta) es un número mixto a (b/c), ENTONCES se convierte en fracción impropia multiplicando la parte entera por el denominador de la parte fraccionaria y sumando el resultado al numerador (a×c)+b. El denominador no varía. 5. SI la fracción resultado de la suma/resta es simplificable ENTONCES se reduce dividiendo numerador y denominador por el máximo común divisor (d):(r/d)/(s/d) 6. SI la fracción resultado de la suma/resta es impropia, ENTONCES se convierte en número mixto dividiendo el numerador entre el denominador, dejando el cociente como parte entera y el resto entre el denominador como parte fraccionaria. 7. SI las dos fracciones b/c, e/f a sumar (restar) tienen el mismo denominador (c=f) ENTONCES se suman/restan los numeradores para obtener el numerador del resultado. El denominador del resultado es igual al denominador de cualquiera de los dos sumandos. (b±e)/c(f) |
|---|

La fiabilidad de la prueba estimada mediante el coeficiente alfa de Cronbach arrojó un valor de 0,884. La media del índice de

dificultad de los ítems (\bar{p}) fue de 0,64 y la media de las correlaciones biserial-puntuales (\bar{r}_{bp}) resultó ser igual a 0,54.

Tabla 2
Reglas de producción requeridas para la resolución de los ítems del test de fracciones

Ítem	Reglas de Producción							
	1	2	3	4	5	6	7	
$2\frac{1}{2} + 3\frac{1}{3}$	1.	1	1	0	0	1	0	1
$\frac{2}{6} - \frac{2}{10}$	2.	0	0	0	1	0	1	1
$\frac{1}{3} + \frac{2}{3}$	3.	1	0	0	0	0	0	1
$\frac{6}{8} + \frac{4}{8}$	4.	1	0	0	0	0	1	1
$\frac{3}{5} - \frac{2}{7}$	5.	0	1	0	0	0	0	1
$3\frac{3}{4} - 2\frac{1}{8}$	6.	1	0	1	0	1	0	1
$3\frac{6}{8} - 2\frac{4}{8}$	7.	1	0	0	0	1	1	1
$3\frac{3}{6} - \frac{4}{12}$	8.	1	0	1	0	1	1	1
$2\frac{1}{3} + 1\frac{5}{6}$	9.	1	0	1	0	1	0	1
$\frac{3}{4} - \frac{1}{8}$	10.	0	0	1	0	0	0	1
$3\frac{3}{5} - 2\frac{1}{5}$	11.	1	0	0	0	1	0	1
$3\frac{4}{6} + 4\frac{2}{12}$	12.	1	0	1	0	1	1	1
$\frac{2}{6} + \frac{2}{8}$	13.	0	0	0	1	0	1	1
$\frac{1}{3} + \frac{5}{6}$	14.	1	0	1	0	0	0	1
$\frac{3}{6} - \frac{4}{12}$	15.	0	0	1	0	0	1	1
$\frac{4}{6} + \frac{2}{12}$	16.	0	0	1	0	0	1	1
$\frac{6}{8} - \frac{4}{8}$	17.	0	0	0	0	0	1	1
$\frac{1}{3} - \frac{2}{7}$	18.	0	1	0	0	0	0	1
$3\frac{1}{3} + 1\frac{2}{7}$	19.	1	1	0	0	1	0	1
$\frac{3}{5} - \frac{1}{5}$	20.	0	0	0	0	0	0	1
$2\frac{6}{8} + 2\frac{4}{8}$	21.	1	0	0	0	1	1	1

Unidimensionalidad

El análisis paramétrico de la unidimensionalidad se lleva a cabo mediante un análisis de componentes principales a partir de la matriz de correlaciones tetracóricas. Se obtiene como resultado un valor propio de 9,79 que explica el 46,60% de la varianza.

Se aplica asimismo el procedimiento DIMTEST para la evaluación de la unidimensionalidad esencial (Nandakumar y Stout, 1993; Stout, 1987). Los resultados se recogen en la tabla 4. La prueba supera el criterio propuesto por Stout aceptándose en consecuencia la hipótesis de unidimensionalidad.

Validación

En la tabla 5 se muestran las matrices C (valores por debajo de la diagonal) y Q (valores por encima de la diagonal). Sus elementos son los valores resultantes de la aplicación de las fórmulas 2 y 3.

Para llevar a cabo el análisis QA, se utilizó la rutina QAP implementada en el software de análisis de redes UCINET IV (Borgatti, Everett y Freeman, 1992).

Los estadísticos descriptivos de las matrices Q y C se muestran en la tabla siguiente.

Los elementos de Q varían entre 0 y 189. La media fue de 43,34 con una desviación estándar de 43,21. Los valores de los elementos de C varían entre 0 y 4 e indican el número de reglas de producción requeridas para resolver cada par de ítems.

El índice gamma de Hubert arrojó un valor de 31,992.

La significación de Γ_{QC} puede establecerse asumiendo bajo la hipótesis nula que las filas y las columnas de las matrices están emparejadas entre sí al azar (Baker y Hubert, 1977). Si se obtuvieran todas las permutaciones posibles entre elementos de filas y columnas, un nivel de significación para el índice observado Γ_{QC} sería simplemente la proporción de índices tanto o más extremos que Γ_{QC} . En la práctica, sin embargo, la significación se establece mediante procedimientos Montecarlo debido a la impracticidad de una prueba exacta a partir de todas las permutaciones posibles (Cliff y Ord, 1981).

La significación de este índice se comprueba mediante 10.000 simulaciones en las que se permutan aleatoriamente las filas y columnas de la matriz C. Entre las 10.000 simulaciones no se obtuvo ningún índice igual o mayor que el observado, por lo que la

Tabla 3
Medias aritméticas y varianzas por cursos

1º E.S.O.		2º E.S.O.		3º E.S.O.		TOTAL	
\bar{X}	S_x^2	\bar{X}	S_x^2	\bar{X}	S_x^2	\bar{X}	S_x^2
13,8	26,1	14,1	32,1	13,0	25,4	13,5	27,2

Tabla 4
Evaluación de la unidimensionalidad esencial

T CONSERVADOR				T POTENTE			
T1	T2	T	p	T1	T2	T	p
1,87104	3,12166	-0,8843	0,81174	2,34268	3,67934	-0,9452	0,82771

Tabla 5
Datos de proximidad, matriz empírica y matriz estructural

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0	20	53	53	68	96	93	98	97	10	97	97	17	39	0	3	2	68	156	0	90
2	0	0	3	3	18	4	45	35	3	2	2	38	98	3	43	46	41	18	20	0	36
3	1	0	0	53	3	42	38	43	44	3	49	49	2	41	0	0	0	3	42	0	41
4	1	1	1	0	1	39	88	77	41	1	44	78	46	37	45	47	45	2	38	0	76
5	1	0	0	0	0	4	1	5	4	11	2	5	13	7	2	3	0	71	71	0	1
6	2	0	1	1	0	0	94	158	154	64	90	154	10	92	67	71	4	6	87	0	87
7	2	1	1	2	0	2	0	133	93	2	95	126	49	32	50	53	49	3	91	0	130
8	2	1	1	2	0	3	3	0	163	63	98	189	45	97	104	110	40	7	95	0	127
9	2	0	1	1	0	3	2	3	0	65	98	158	8	99	64	68	1	4	95	0	91
10	0	0	0	0	0	1	0	1	1	0	2	62	6	64	59	62	0	8	10	0	1
11	2	0	1	1	0	2	2	2	2	0	0	101	3	38	1	1	1	4	95	0	95
12	2	1	1	2	0	3	3	4	3	1	2	0	43	97	99	104	37	7	92	0	122
13	0	2	0	1	0	0	1	1	0	0	0	1	0	8	52	54	47	11	17	0	42
14	1	0	1	1	0	2	1	2	2	1	1	2	0	0	60	63	0	6	39	0	34
15	1	0	0	1	0	2	1	2	2	1	1	2	0	2	0	119	47	2	2	0	40
16	0	1	0	1	0	1	1	2	1	1	0	2	1	1	1	0	51	3	3	0	42
17	0	1	0	1	0	0	1	1	0	0	0	1	1	0	0	1	0	0	1	0	43
18	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	71	0	3
19	3	0	0	1	1	2	2	2	2	0	2	2	0	1	1	0	0	1	0	0	93
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	2	1	1	2	0	2	3	3	2	0	2	3	1	1	1	1	1	0	2	0	0

asociación entre las dos matrices es altamente significativa ($p < 0,0001$).

Evaluación del cumplimiento de los supuestos del modelo

La adecuación de un modelo logístico lineal exige en primer lugar la aplicabilidad del modelo de Rasch a los datos. Con el fin de garantizarlo se realizan una serie de comprobaciones previas que, siguiendo las pautas de (Hambleton, Swaminathan, y Rogers, 1991), examinarán la validez de los supuestos así como el cumplimiento de las propiedades esperadas.

La comparación de la distribución observada del número de aciertos en los 4 ítems más difíciles tomando el grupo de sujetos por debajo del percentil 20, con la distribución de frecuencias esperadas

en base a la distribución binomial ($n = 4, \pi = 0,33$) arroja un valor del estadístico Ji cuadrado de 116,83 ($p < 0,000$) por lo que se rechaza la hipótesis de elección aleatoria de las alternativas de respuesta.

El grado de asociación lineal entre las estimaciones obtenidas en los dos subgrupos de sujetos (habilidad alta/habilidad baja) utilizando la mediana como punto de corte se traduce en un coeficiente de correlación de 0,91 resultado que en principio parece apoyar el cumplimiento de invarianza de los parámetros de dificultad (Gráfico 1).

Por último, el rango de la distribución de los índices de correlación biserial puntual es de 0,35 siendo su coeficiente de variación de 0,19.

Una vez ajustado el modelo de Rasch y estimados los parámetros de dificultad se obtuvieron valores de Ji cuadrado significativos para 7 ítems ($\alpha = 0,01$).

Tabla 6
Estadísticos descriptivos de las matrices Q y C

MATRIZ Q	MATRIZ C
$\bar{X} = 43,34$	$\bar{X} = 0,88$
$S = 43,21$	$S = 0,93$
Min = 0,00	Mín = 0,00
Máx = 189,00	Máx = 4,00

Tabla 7
Estimación de los parámetros básicos del LLTM

Componente	$\hat{\alpha}$	Error estándar	Valor z
1.	0,1173	0,0719	1,6322
2.	0,7299	0,0484	15,0713
3.	1,2292	0,0808	15,2123
4.	1,0990	0,0647	16,9952
5.	0,7028	0,0488	14,3982
6.	0,7591	0,0705	10,7699

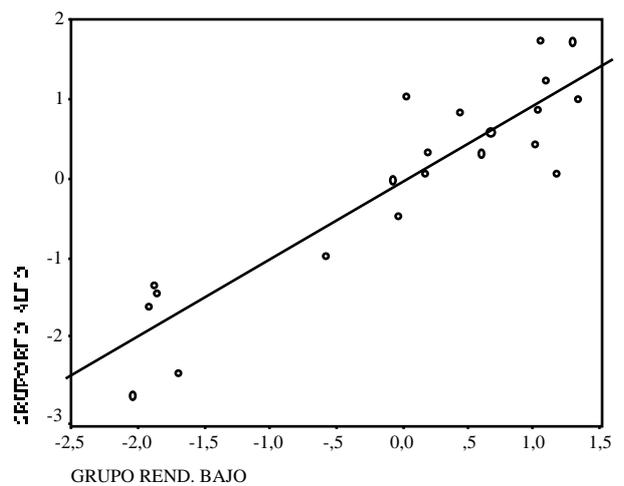


Gráfico 1. Relación lineal entre las estimaciones de dificultad en el modelo de Rasch basados en dos grupos de sujetos en función de su rendimiento

Aplicación del modelo LLTM

El modelo logístico lineal es desde el punto de vista formal más general que el modelo de Rasch, reduciéndose éste a un caso particular de aquél en el que la matriz de pesos es una matriz identidad. Sin embargo desde el punto de vista de su aplicabilidad es un modelo más «exigente» estableciendo ciertas restricciones lineales a los parámetros de los ítems, siendo éstas las que permiten la especificación de las contribuciones de cada operación a la dificultad del ítem.

Las estimaciones de los parámetros básicos para cada componente ($\hat{\delta}$), sus respectivos errores estándares y valores z obtenidos mediante el programa LPCM-Win 1.0 (Fisher y Ponocny-Seliger, 1998) aparecen en la tabla 7. Es preciso señalar en este punto que el componente de dificultad asociado con la regla de producción referida a la suma/resta de fracciones con el mismo denominador (regla 7) está presente en todos los ítems de la prueba por lo que en la práctica son seis las columnas de la matriz de pesos.

Los componentes son todos significativos, excepto el primero que refleja la dificultad más baja. Los errores estándares son de escasa magnitud. La media de las estimaciones de los parámetros básicos es de 0,77 con una desviación típica de 0,39. Los valores oscilan entre 0,1173 y 1,2292. Este último valor que refleja el componente más difícil, corresponde a la regla de producción implicada en la adición/sustracción de fracciones en el caso de denominadores distintos, no primos entre sí y con mínimo común múltiplo diferente a ambos. La complejidad de la regla, que exige la descomposición en factores primos, puede ser la razón de esta dificultad.

El cuarto componente, relativo a la transformación de un número mixto en una fracción impropia, muestra también una dificultad relativamente alta ($\hat{\alpha} = 1,0990$) que curiosamente difiere notablemente del grado de dificultad correspondiente a la regla sexta ($\hat{\alpha} = 0,7591$) la cual requiere también del (re)conocimiento de los números mixtos pero implicando la conversión inversa. Una posible explicación sería la presencia de un fenómeno de interferencia o confusión en algunos sujetos, entre un número mixto y la operación de multiplicación entre un número entero y una fracción. La ambigüedad de la notación favorecería la probabilidad de aparición de este error (Tatsuoka et al., 1984).

Una vez obtenidas las estimaciones, el ajuste del modelo logístico lineal se pone a prueba en primer lugar por el grado de relación lineal entre los parámetros estimados sobre la base del modelo de Rasch y los parámetros estimados sobre la base del modelo LLTM. El coeficiente de correlación entre ambas series es de 0,9028. En segundo lugar, el test estadístico de Andersen (Andersen, 1973) permitirá el contraste entre la hipótesis nula de ajuste al modelo logístico lineal, frente a la hipótesis alterna de ajuste al modelo de Rasch. El valor χ^2 del estadístico obtenido por la fórmula para nuestros datos es de 541,6 siendo por lo tanto significativo ($\chi_{14}^2 = 23,6735$, $p < 0,05$) y en consecuencia se rechazará H_0 concluyendo en aceptar el modelo de Rasch sin que se proporcione justificación estadística para utilizar el LLTM en lugar de aquél.

Discusión

Los resultados de este trabajo muestran una doble vertiente; Por un lado, el estudio de validación realizado a través del análisis de asignación cuadrática ofrece resultados altamente satisfactorios; La correspondencia entre las dos matrices, empírica y teóri-

ca, sugieren que los sujetos emplean las operaciones definidas en el marco del LLTM.

Sin embargo los resultados referidos al estricto ajuste estadístico del modelo distan de ser concluyentes. Aun existiendo un alto grado de correlación entre las dificultades pronosticadas por el modelo de Rasch y las obtenidas por el LLTM sin embargo el estadístico J_i cuadrado de la razón de verosimilitud no avala un buen ajuste. Hagamos notar no obstante que tal y como reconoce el mismo Fisher (1995) en la gran parte de las aplicaciones se obtienen resultados significativos de J_i cuadrado conducentes al rechazo del modelo (Medina Díaz, 1993; Real, Olea, Ponsoda, Revuelta, y Abad, 1999). Basar la adecuación del modelo únicamente en dicho índice puede resultar arriesgado, de ahí el valor de procedimientos complementarios como el desarrollado en este trabajo, encaminados a validar la matriz estructural de componentes.

Por otro lado el ajuste de los datos al modelo de Rasch no ha resultado óptimo. Aunque los resultados de las pruebas estadísticas encaminadas a comprobar el supuesto de ausencia de azar y la propiedad de invarianza han dado resultados positivos, no se puede obviar el relativamente alto porcentaje de ítems desajustados, aunque habría que considerar en este punto la inestabilidad del índice J_i cuadrado señalada por diversos autores (Embretson y Reise, 2000; Mislevy y Bock, 1990; Suárez y Martínez, 2000).

Un aspecto que consideramos interesante para tratar en futuros trabajos es el de la comparación de diferentes modos de cuantificación de los pesos; En este trabajo se ha tenido en cuenta la presencia/ausencia de una regla u operación en la resolución de un ítem. Un procedimiento alternativo consistiría en la consideración de la *frecuencia* con que la aplicación de dicha regla tiene lugar. Por otro lado se podrían considerar también operativizaciones diferentes de la similitud estructural entre los ítems a la utilizada aquí (número de reglas comunes).

Por último aventuramos la hipótesis de que un modelo menos restrictivo que el de Rasch, que aúne la inclusión de parámetros adicionales con la posibilidad de descomposición de la dificultad en términos de las operaciones cognitivas podría adecuarse mejor a los datos (Embretson, 1999). Recordemos en este punto que la variabilidad de los parámetros de discriminación apuntan a un modelo que incorpore el parámetro de discriminación como alternativa adecuada.

En cualquier caso los resultados del estudio muestran que el proceso por el cual los sujetos resuelven la tarea de adición/sustracción de fracciones y números mixtos puede representarse válidamente. Ahora bien, el modo de resolución no es independiente de los algoritmos y métodos enseñados efectivamente en el aula, por lo tanto, somos cautos ante la generalización de resultados sin un análisis previo de los procedimientos instruccionales.

Desde el punto de vista didáctico, teniendo en cuenta que cada regla representa un fragmento significativo del proceso de solución total del problema, la información acerca de las dificultades relativas de las reglas permitirá identificar las áreas que requieren una mayor atención y refuerzo. Si bien es cierto que la competencia en el dominio de las fracciones es de carácter multidimensional, conceptual, factual y algorítmico o procedimental (English y Halford, 1995), es indudable que a esta última dimensión se le concede gran atención en el desarrollo curricular.

Para finalizar señalemos que, además de la cuantificación de las dificultades relativas de cada regla obtenidas mediante el modelo podrían utilizarse para desarrollar ítems con dificultades

predeterminadas y en último término, desarrollando el algoritmo generativo adecuado, podrían incorporarse a un sistema de Generación Automática de Items (GAI) (ver Revuelta y Ponsoda, 1999).

Agradecimientos

Este trabajo ha sido financiado por la Universidad del País Vasco. UPV (1/UPV/EHU 00109.231-HA-7852/2000).

Referencias

- Andersen, E.B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.
- Anderson, J.R. (1976). *Language, Memory and Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Baker, F.B. y Hubert, L.J. (1977). Inference procedures for ordering theory. *Journal of Educational Statistics*, 2(3), 217-233.
- Borgatti, S.P., Everett, M.G. y Freeman, L.C. (1992). *Ucinet IV Version 1.35*. Columbia, SC: Analytic Technologies.
- Butter, R. y De Boeck, P. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika*, 63(1), 47-63.
- Cliff, A.D. y Ord, J.K. (1981). *Spatial processes: Models and applications*. London: Pion.
- Embretson, S.E. (1983). Construct validity: construct representation versus nomothetic span. *Psychological Bulletin*, 93(1), 179-197.
- Embretson, S.E. (1984). A General Latent Trait Model for response processes. *Psychometrika*, 49(2), 175-186.
- Embretson, S.E. (1985). Introduction to the problem of test design. En S.E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics*. Orlando, Florida: Academic Press, Inc.
- Embretson, S.E., y Wetzel, D. (1987). Component latent trait models for paragraph comprehension tests. *Applied psychological measurement*, 11, 175-193.
- Embretson, S.E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56, 495-515.
- Embretson, S.E. (1994). Applications of Cognitive Design Systems to Test Development. En C.R. Reynolds (Ed.), *Cognitive Assessment; A Multidisciplinary Perspective*. N.York: Plenum Press.
- Embretson, S.E. (1995a). A measurement model for linking individual learning to processes and knowledge: application to mathematical reasoning. *Journal of educational measurement*, 32(3), 277-294.
- Embretson, S.E. (1995b). The role of working memory capacity and general control processes in intelligence. *Intelligence*, 20, 169-189.
- Embretson, S.E. (1999). Generating items during tests: Psychometric issues and models. *Psychometrika*, 64(4), 407-433.
- Embretson, S.E. y Reise, S.P. (2000). IRT applications in cognitive and developmental assessment. En S.E. Embretson y S.P. Reise (Eds.), *Item response theory for psychologists*. London: Lawrence Erlbaum.
- English, L.D.H. y Halford, G.S. (1995). *Mathematics Education. Models And Processes*. Mahwah, New Jersey, Lawrence Erlbaum Associates, Publishers.
- Fisher, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta psychologica*, 37, 359-374.
- Fisher, G.H. (1977). Linear logistic trait models: Theory and application. En H. Spada y W.F. Kempf (Eds.), *Structural models of thinking and learning* (pp. 203-225). Bern: Huber.
- Fisher, G.H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48, 3-26.
- Fisher, G.H. (1995). The Linear Logistic Test Model. En G.H. Fisher e I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications*. New York: Springer-Verlag.
- Fisher, G.H. y Ponocny-Seliger, E. (1998). *Structural Rasch modeling: Handbook of the usage of LPCM-WIN 1.0*. Groningen: ProGAMMA.
- Frederiksen, N., Mislevy, R. y Béjar, I.I. (1993). *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hambleton, R.K., Swaminathan, H. y Rogers, H.J. (1991). *Fundamentals of item response theory*: Sage Publications.
- Homke, L.F. y Habon, M.W. (1986). Rule-based item bank construction and evaluation within the linear logistic framework. *Applied psychological measurement*, 10(4), 369-380.
- Hubert, L. y Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29, 190-241.
- Hubert, L., Golledge, R.G., Costanzo, C.M., Gale, N. y Halperin, W.C. (1984). Nonparametric tests for directional data. En G. Bahrenberg, M.M. Fisher y Nijkamp (Eds.), *Recent developments in spatial data analysis* (pp. 171-189). Brookfield VT: Brower.
- Hubert, L.J. (1985). Combinatorial data analysis: Association and partial association. *Psychometrika*, 50(4), 449-467.
- Maris, E.M. (1995). Psychometric latent response models. *Psychometrika*, 60(4), 523-547.
- Medina Díaz, M. (1993). Analysis of cognitive structure using the linear logistic test model and quadratic assignment. *Applied psychological measurement*, 17(2), 117-130.
- Mislevy, R.J. y Bock, R.D. (1990). *BILOG 3: Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville, IN: Scientific Software.
- Mislevy, R.J. y Verhelst, N.D. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55, 195-215.
- Nandakumar, R. y Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics*, 18, 41-68.
- Newell, A. y Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Pereda, L. (1975). *Matemáticas. Desarrollos didácticos*. Erein.
- Prieto, G. y Delgado, A.R. (2000). Utilidad y representación en la psicometría actual. *Revista de Metodología de las Ciencias del Comportamiento*, 2(2), 111-127.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.
- 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 exponencial. *Psicológica*, 20, 121-134.
- Revuelta, J. y Ponsoda, V. (1999). Generación automática de ítems. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests adaptativos Informatizados: Fundamentos y aplicaciones*. Madrid: Pirámide.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282.
- Simon, H.A. (1978). Information-processing theory of human problem solving. En W.K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 271-295). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Smith, R.M., Kramer, G.A. y Kubiak, A.T. (1992). Components of difficulty in spatial ability test items. *Objective Measurement: Theory into Practice, Vol. 1* (pp. 157-174).
- Spada, H. y May, R. (1982). The linear logistic test model and its application in educational research. En D. Spearrit (Ed.), *The improvement of measurement in Education and Psychology* (pp. 67-84). Hawthorne, Victoria: Australian council for educational research.
- Spada, H. y McGaw, B. (1985). The assessment of learning effects with linear logistic test models. En S. Embretson (Ed.), *Test design: developments in psychology and psychometrics*. Orlando, Florida: Academic Press, Inc.
- Stout, W.F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617.
- Suárez, J.C. y Martínez, R. (2000). Análisis del funcionamiento de índices de ajuste al ítem ante la violación del supuesto de curvas paralelas del modelo de Rasch. *Revista de Metodología de las ciencias del comportamiento*, 2(1), 49-62.

- Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Tatsuoka, K.K., Birenbaum, M., Yamamoto, K. y Dowd, B. (1984). *Analysis of errors in fraction addition and subtraction problems* (NIE-G-81-0002). Urbana - Champaign: University of Illinois.
- Whitely, S.E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45(4), 175-186.
- Yamamoto, K. y Gitomer, D.H. (1993). Application of a HYBRID model to a test of cognitive skill representation. En N. Frederiksen, R. Mislevy e y I. BÉjar (Eds.), *Test theory for a new generation of tests* (pp. 275-295). Hillsdale, NJ, Lawrence Erlbaum Associates, Inc.