Revista Electrónica de Metodología Aplicada
2006, Vol. 11 nº 1, pp. 1-15

Modelo de doble monotonía fuerte para ítems politómicos aplicado a una escala de autoconcepto social. Propiedades formales y métricas[1]

Paula Elosua Oliden[2]
Facultad de Psicología. Universidad del País Vasco.

RESUMEN
En este trabajo se describen los supuestos formales y las características métricas del modelo no-paramétrico fuerte de doble monotonía en su versión politómica. Se presentan distintos procedimiento para su evaluación empírica. Se ajusta el modelo a una escala politómica de autoconcepto social, y se presentan sus ventajas y principales carencias.
Palabras clave: Modelo no-paramétrico de respuesta al ítem, Modelo fuerte de doble monotonía, Modelo politómico, Mokken, Orden estocástico.

ABSTRACT
This paper describes the formal and metric characteristics of one non-parametric model of item response theory: The strong double monotonicity model for polytomous items. The methods for empiric evaluating are presented. In order to illustrate this model, it is applied to one polytomous scale about social selfconcept. Finally the advantages and problems with this model are presented.
Keywords: Non-parametric item response model. Strong double monotonicity. Polytomous model. Mokken. Stochastic ordering.

---

[1] Trabajo financiado por la Dirección General de Investigación del Ministerio de Educación y Ciencia BSO2002-00490
[2] FACULTAD DE PSICOLOGÍA
AVDA. TOLOSA, 70
20009 SAN SEBASTIÁN
E-mail. paula.elosua@ehu.es

1.- Introducción

Desde los primeros trabajos de Thurstone (1927) y Guttman (1950) en el campo de la medición de actitudes el supuesto básico asumido por todos los modelos de escalamiento ha sido la existencia de una escala psicológica latente (θ) del constructo medido. La estimación de los valores o puntos que representan a las personas o a los objetos sobre ésta se apoya en algún tipo de función que la relaciona con la escala empírica (test, escala, cuestionario...) utilizada como instrumento de medida. Las puntuaciones en la escala empírica habitualmente, se corresponden con una suma directa o ponderada (X+) de las puntuaciones obtenidas en cada uno de sus componentes o ítems (X_i). Esta correspondencia X+⇔ θ comporta en cualquier caso, el cumplimiento de condiciones métricas que garantizarán las inferencias relacionadas con el orden bien de los sujetos bien de de los ítems. Las propiedades métricas que rigen este ordenamiento son las propiedades de orden :

-*Ordenamiento estocástico de la variable manifiesta.* OEM. El orden de los sujetos sobre la variable latente (θ) produce un ordenamiento estocasticamente correcto de los sujetos sobre la variable manifiesta. Para dos sujetos A y B con niveles en la variable latente θ_A y θ_B, tal que $\theta_A<\theta_B$, OEM establece que para cualquier puntuación manifiesta x, siendo X+ la puntuación empírica total

$$P\left(X+\geq x\middle|\theta_A\right)\leq P\left(X+\geq x\middle|\theta_B\right) \qquad (1)$$

- *Ordenamiento estocástico de la variable latente* OEL. El orden de los sujetos sobre la variable manifiesta (X+) produce un ordenamiento estocasticamente correcto de los sujetos sobre la variable latente. Para un valor constante en la variable latente (s) y para dos valores manifiestos x_1 y x_2, tal que $x_1<x_2$,

$$P\left(\theta\geq s\middle|X+=x_1\right)\leq P\left(\theta\geq s\middle|X+=x_2\right) \qquad (2)$$

- *Ordenamiento invariante de los ítems* (OII). El orden de los ítems en la escala latente es el mismo para todos los valores de theta. OII define la esperanza condicional de la puntuación empírica, $E(X_i|\theta)$ y confirma su ordenamiento a lo largo del continuo subyacente,

$$E\left(X_1|\theta\right)\leq E\left(X_2|\theta\right)\leq\ldots\leq E\left(X_k|\theta\right) \qquad (3)$$

El ordenamiento de las puntuaciones y el ordenamiento de los sujetos queda implícito en *la razón de verosimilitud monótona* (RVM) (Hemker, Sijtsma, Molenaar, Junker, 1996), de modo que para K ítems y dos puntuaciones manifiestas tal que $0\leq x_1<x_2\leq X_{max}$,

$$g\left(x_1,x_2|\theta\right)=\frac{P\left(X_+=x_1|\theta\right)}{P\left(X_+=x_2|\theta\right)} \qquad (4)$$

es una función no-decreciente de theta. La RVM es una propiedad más general que OEM u OEL, dado que aquella implica estas pero estas no implican la RVM.

Sobre el resto de las propiedades de orden, OEM, OEL e IIO no puede establecerse ningún tipo de relación de implicación.

La presencia de estas propiedades métricas queda sujeta a la verificación de características formales que no están presenten en todos los modelos psicométricos utilizados para la medición de atributos psicológicos. Los supuestos mínimos que caracterizan los modelos de respuesta al ítem, unidimensionalidad, independencia local y monotonía, garantizarían sólo el ordenamiento estocástico de la variable manifiesta (OEM). El cumplimiento del resto de las propiedades están asociadas con requisitos formales que sólo se verifican totalmente para el modelo dicotómico de Rasch y sus derivados politómicos: Modelo de Crédito Parcial de Masters (1982) y de Andrich (1978). Se trata de modelos muy restrictivos y por tanto difíciles de aplicar en condiciones empíricas. Frentes a ellos y frente al resto de modelos de teoría de respuesta al ítem paramétricos el investigador dispone de modelos más flexibles y menos exigentes, los modelos de respuesta al ítem no-paramétricos, que también poseen y garantizan propiedades de orden .

El objetivo de este trabajo es mostrar las propiedades formales y métricas de un modelo politómico no-paramétrico de respuesta al ítem, y ejemplificarlo por medio de un estudio empírico basado en una escala de autoconcepto social (Musitu, García y Gutiérrez, 1997). Se trata del modelo fuerte de doble monotonía (Sijtsma y Hemker, 1998) que es una extensión y generalización del modelo dicotómico de Mokken (Mokken, 1971, 1997; Mokken y Lewis, 1982). El modelo de Mokken tiene ya una tradición de más de 30 anos y ha sido utilizado con éxito en campos como la construcción de tests (Sijtsma y Molenaar, 2002), las ciencias políticas (van Schuur, 2003) y en general en aquellas situaciones en que los reactivos pueden considerarse acumulativos, y las características de los datos no se ajustan a modelos paramétricos que por su naturaleza son más restrictivos.

2.- Modelo fuerte de doble monotonía para ítems politómicos

2.1.- Caracterización de los modelos politómicos de repuesta graduada

Los formatos politómicos de respuesta presentan para cada ítem (*i*) *m* categorías de respuesta que pueden ordenarse y codificarse como 0, 1,..., *m*-1. Un ítem politómico divide el continuo latente en *m* categorías ordenadas separadas por *m-1* "*pasos entre categorías*" (Figura 1). Esta representación permite distinguir entre la puntuación obtenida en el ítem (X_i) y los valores entre las opciones de respuesta, también llamados pasos. Los pasos entre categorías pueden representarse como variables dicotómicas que adoptan dos valores 0 y 1 (Y_{ig}). En función de cómo se defina esta dicotomía, es decir, la relación entre las categorías, es posible diferenciar entre modelos de respuesta graduada o acumulativos, modelos secuenciales y modelos de crédito parcial o adyacentes (Glas, Scheerens y Thomas, 2003). El modelo que nos ocupa se encuentra en la primera categoría.

La relación entre X_i y Y_{ig} viene dada por el total de pasos (g) superados en el ítem *i*; Por ejemplo para el caso de 1-2-3 opciones de respuesta (*m*=3) tendríamos dos pasos entre opciones (Yi_0, Yi_1). El paso de ítem Y_{i0} se refiere a la variable que enfrenta la opción de respuesta 1 (recodificada como 0) con las opciones 2 y 3 (recodificadas como 1 y 2). El paso Y_{i1} se considera un ítem dicotómico en el que la dicotomía esta representada por las opciones 12 (valores 01) con la opción 3 (valor 2).

Una persona que elige la opción 3 de respuesta, ha superado dos pasos entre opciones por lo que se le asignara la puntuación 2. Se trata de un tipo de escala “acumulada” en el que la acumulación se lleva a cabo entre la “dificultad” de los pasos de los ítems. Su representante paramétrico se correspondería con la familia de modelos de respuesta graduada (Samejima, 1969, 1997).

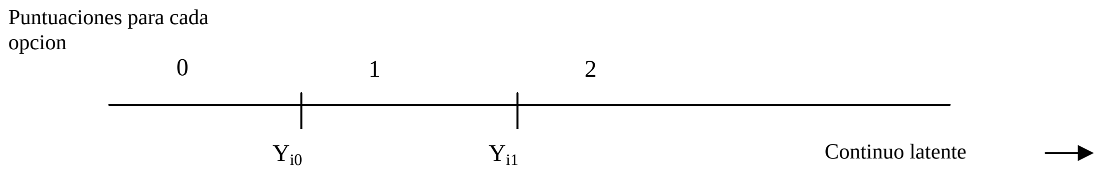

Figura 1 Posición sobre el continuo latente de los 2 pasos entre opciones correspondientes a un ítem con 3 categorías de respuesta.

Este modo de representación de los ítems politómicos permite tratar a los modelos como si fueran modelos dicotómicos aplicados a (*m*)*n* pasos donde *n* es el número de ítems que componen el test. Su caracterización formal se define a través de la función de respuesta al paso del ítem (ISRF; *item sep response function*). La ISRF describe la relación entre la probabilidad de superar una opción (Y_{ig}=1) y el rasgo latente, o equivalentemente entre la probabilidad de que la puntuación del ítem sea mayor que la opción *g* ($X_i \geq g$), y el rasgo latente. $P\left(Y_{ig} = 1|q\right) = P\left(X_i\ {}^3\ \ g|q\right)$.

2.2.- Supuestos del modelo

A los supuestos básicos de unidimensionalidad, independencia local y monotonía, hay que añadir dos condiciones más: la doble monotonía, y la doble monotonía fuerte.

-*Monotonía:* (M) Las funciones de respuesta entre categorías para cada uno de los ítems son monótonas no-decrecientes en θ (Figura 2a). M asegura el ordenamiento de lo sujetos a lo largo del continuo de habilidad. Así para cualquier par de valores θ tal que $\theta_v < \theta_w$, y para todo ítem (*i*) y opción de respuesta (*g*),

$$P\left(X_i \geq g\,|\theta_v\right) \leq P\left(X_i \geq g\,|\theta_w\right) \qquad (5)$$

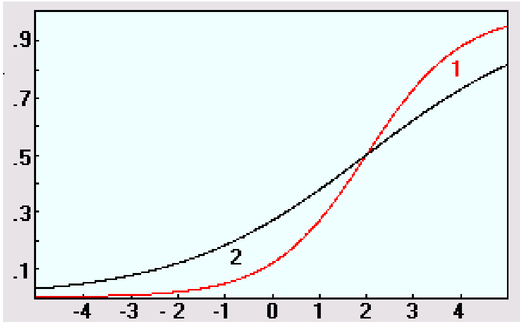

Figura 2a. Homogeneidad monótona

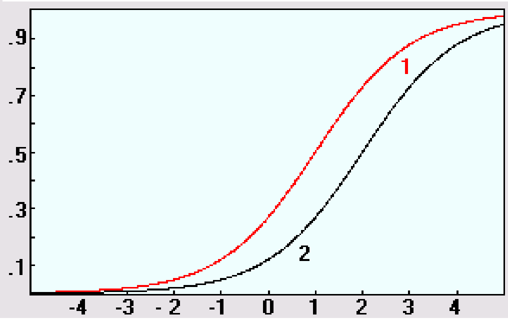

Figura 2b. Doble monotonía

- *Doble monotonía*. (DM) Las funciones de respuesta entre *categorías de ítems diferentes* no se cruzan, es decir, tienen un ordenamiento invariante para todos los sujetos en todos los niveles de habilidad (Figura 2b). Para dos ítems *i* y *j* y dos pasos entre opciones g_1 y g_2 tal que $g_1<g_2$ se cumple para todo θ

$$P\left(X_i \geq g_1 \middle| \theta\right) \leq P\left(X_j \geq g_2 \middle| \theta\right) \tag{6}$$

En este ordenamiento invariante de las funciones de respuesta entre pasos no está implícito el ordenamiento de los ítems. En el modelo DM es posible encontrar para dos ítems *i* y *j* que $P(X_i>=1|\theta)< P(X_j>=1|\theta)$ aunque $P(X_i>=2|\theta)> P(X_j>=2|\theta)$

- *Doble monotonía fuerte*. (DMF) Añade a la condición de doble monotonía una restricción por pares que asegura el *ordenamiento invariante de los ítems* (Sijtsma y Hemker, 1998) de modo que ante cualquier paso entre opciones de respuesta (g) y dos ítems *i*, *j*:

$$P\left(X_i \geq g \middle| \theta\right) \leq P\left(X_j \geq g \middle| \theta\right) \tag{7}$$

Desde una perspectiva no-paramétrica el orden invariante de los ítems tiene una estrecha relación con el concepto de funcionamiento diferencial del ítem. Si los ítems se relacionan con el constructo a medir del mismo modo en dos muestras (G1 y G2), entonces el orden de los ítems sobre el continuo latente habrá de ser el mismo para las dos muestras; será invariante. Formalmente podríamos decir que la esperanza matemática de una puntuación empírica condicionada sobre el nivel de habilidad (θ) (ecuación 3) es independiente de la pertenencia al grupo.

$$E\left(X = x \middle| \theta, G_1\right) = E\left(X = x \middle| \theta, G_2\right) \tag{8}$$

El cumplimiento de los supuestos del modelo no-paramétrico fuerte de doble monotonía para datos politómicos garantiza las propiedades métricas de ordenamiento invariante de los ítems y de ordenamiento estocástico manifiesto pero no el ordenamiento estocástico latente (Grayson, 1988; Hemker y Sijtsma, 1999; van der Ark, 2001). Es decir cuando los datos cumplan los supuestos descritos será posible ordenar a los ítems sobre el continuo latente (Propiedad OII), será posible ordenar a los sujetos sobre la variable manifiesta en función de sus valores en la variable latente (OEM) pero no esta garantizado el ordenamiento en la variable latente en función de los valores obtenidos por los sujetos en la escala empírica (OEL).

2.3.- Evaluación del modelo

La evaluación de los supuestos de ese modelo se basa en la verificación de las propiedades empíricas que se derivan de ellos. Si se cumple la monotonía podremos encontrar estas características en los datos: a) todos los pares de ítems tendrán correlaciones no-negativas para todos los subgrupos de sujetos que difieren en el mismo rasgo (Mokken, 1971); b) todo par de ítems tendrá una asociación condicional (CA) en cualquier grupo de personas con una particular puntuación empírica (X+) (Rosenbaum, 1984); c) para cualquier ítem *i* la proporción de personas que lo respondan correctamente será no-decreciente sobre los grupos de puntuación creciente (X+), donde X+ se estima sobre el resto de *n*-1 ítems; finalmente d) la puntuación total (X+) tendrá una regresión monótona no-decreciente sobre la habilidad (Lord y Novick, 1968), lo que implica una correlación positiva entre ambos. La doble monotonía por su parte implicaría que el orden de las dificultades de los pasos entre ítems es independiente de la distribución de habilidad; esta propiedad aseguraría el ordenamiento invariante de las funciones de respuesta entre pasos. Este ordenamiento entre pasos sólo podría generalizarse al ordenamiento entre ítems si se cumple la doble monotonía fuerte.

- *Coeficiente de escalabilidad*. La evaluación empírica de M puede llevarse a cabo a través de una adaptación del coeficiente de escalabilidad de Loevinger (1947, 1948) que Mokken (1971) utilizó para las escalas dicotómicas. Este coeficiente se define en términos de errores de una escala Guttman que compara un patrón de respuesta observado con el patrón teórico que debiera seguir una escala acumulada.

Para dos ítems *i* y *k* ($i<k$) se operacionaliza del siguiente modo: $H_{ik}=1-(e_{ik}/e_{ik}^{(0)})$ donde e_{ik} y $e_{ik}^{(0)}$ son las probabilidades de errores observados y esperados bajo el modelo de independencia marginal, en cuyo caso H_{ik} sería 0. En la hipotética situación en que los patrones teóricos y observados coincidieran el coeficiente de escalabilidad sería 1. El coeficiente de escalabilidad para un ítem (*i*) podría obtenerse con respecto al resto de *n*-1 ítems como una combinación lineal del total de H obtenidos.

La adaptación de este coeficiente a una escala politómica fue propuesta por Molenaar (1991, 1997; Sijtsma y Molenaar, 2002). El nuevo coeficiente pondera los errores en los patrones de respuesta observados respecto a los patrones teóricos. La ponderación sobre un par de ítems se lleva a cabo teniendo en cuenta el número de pasos entre opciones involucrados en la resolución de esos ítems. El valor de H ponderado iguala la razón entre la correlación observada entre dos ítems y la correlación máxima obtenida a partir de sus frecuencias marginales,

$$H_{ij} = 1 - \frac{\sum_i \sum_j w_{ij} n_{oij}}{\sum_i \sum_j w_{ij} n_{eij}} \qquad (9)$$

donde w_{ij} es el coeficiente ponderador o numero de errores en el patrón ij
n_{oij} es el numero de respuestas observadas en el patrón de respuesta ij
n_{eij} es el numero de respuestas esperadas en el patrón ij bajo el modelo de independencia

Para la evaluación de su significatividad Molenaar y Sijstma (2000) utilizan un test estadístico contra la hipótesis nula de H=0. Sin embargo, dado que en condiciones empíricas este test siempre resulta significativo, es habitual valorar la escalabilidad de un ítem respecto al punto de corte de 0,30 (Mokken, 1971).

- *Evaluación de la homogeneidad monótona*: Puesto que la proporción de personas que superan una opción de respuesta de un ítem es no-decreciente sobre los grupos de puntuación creciente (X+) para comprobar esta propiedad bastaría con formar grupos de sujetos por niveles de puntuación manifiesta (X+), y comprobar que el porcentaje de personas que superan una opción se incrementa a medida que se incrementa ésta.

- *Evaluación de la doble monotonía*: La evaluación de la doble monotonía se lleva a cabo definiendo proporciones univariadas (π_{ig}^{+}) y bivariadas ($\pi_{ig,jr}^{+}$)de respuestas entre categorías. Con la información obtenida se forman matrices empíricas (P(++)) de orden $n(m-1)\times n(m-1)$ con elementos $\pi_{ig,ir}^{+}$. Las filas y las columnas se ordenan crecientemente en función de los marginales de las proporciones π^{+}_{ig} . Dado este ordenamiento, las celdas tanto de las filas como de las columnas deben ser monótonamente no-decrecientes si las ISRF son invariantemente ordenadas en theta.

- *Evaluación de la doble monotonía fuerte*: Los análisis llevados a cabo para estudiar la condición de doble monotonía fuerte son similares a los llevados a cabo en la evaluación de la doble monotonía pero en lugar de comparar las proporciones de respuestas referidas a los pasos entre ítems las comparaciones se llevan a cabo sobre las medias aritméticas obtenidas en los ítems.

3.- Método

3.1.- Participantes

La muestra está formada por 1269 estudiantes distribuidos entre el 6º curso de Educación Primaria, Educación Secundaria y Bachiller. 604 son chicos y 665 son chicas, que se reparten por niveles educativos como sigue: N_{EP6}=110, N_{ESO} = 831, N_{BA} = 328. La selección de los participantes ha transcurrido en dos fases. En una primera etapa, siguiendo un criterio de doble estratificación en función del tipo centro de enseñanza público/privado/concertado y rural/urbano, se han concretado los centros que podrían participar en el estudio. En una segunda fase, fueron criterios circunstanciales los que determinaron los centros participantes.

	CHICOS	CHICAS	Total
Primaria 6	49	61	110
ESO1	93	96	189
ESO2	97	125	222
ESO3	104	111	215
ESO4	105	100	205
BACHILLER1	84	103	187
BACHILLER2	72	69	141
Total	604	665	1269

Tabla 1. Distribución de la muestra por sexo y nivel educativo

3.2.- Instrumento

La escala de autoconcepto social forma parte del cuestionario AFA-A para la medida multidimensional del autoconcepto (Musitsu, Garcia y Gutierrez, 1997). Se trata de una escala de rendimiento típico compuesto por 5 ítems de respuesta ordenada (Siempre, Algunas Veces, Nunca). Contamos con 5 ítems politómicos (n=5) con 3 categorías de respuesta (m=3) y por tanto con 2 funciones de respuesta entre opciones o pasos. La escala total podría representarse por 10 ítems dicotómicos ($n\times(m$-1)).

3.3.- Análisis

Como paso previo al análisis de datos se invierte la escala de respuesta del ítem 1.

Los análisis efectuados comienzan con un estudio descriptivo del comportamiento de cada uno de los ítems y de cada una de sus alternativas de respuesta.

Se estudia el carácter unidimensional de la escala analizada y una vez comprobado que el requisito de unidimensionalidad está presente en los datos, se evalúan las propiedades de monotonía, doble monotonía, doble monotonía fuerte y orden invariante de los ítems. El software empleado en estos análisis es MSP5 (Molenaar y Sijtsma, 2000).

3.4.- Resultados

3.4.1.- Descriptivos y unidimensionalidad.

La media aritmética de la escala es 12,82 y su desviación típica 1,7. El coeficiente de consistencia interna es 0,63. La tabla 2 muestra los índices descriptivos clásicos de los ítems (media aritmética, desviación estándar e índice de discriminación corregido) así como la distribución de las respuestas por categorías.

Item	Media aritmética	Desviación Estándar	Índice discriminación	Respuestas por categoría		
				Nunca	Algunas veces	Siempre
1 Es difícil para mi mantener los amigos	2,72	0,48	0,35	951	296	22
2 Me gusta mi forma de ser	2,45	0,59	0,33	63	557	649
3 Tengo muchos amigos	2,77	0,46	0,44	24	231	1015
4 Soy un chico alegre	2,57	0,54	0,40	30	485	755
5 Consigo fácilmente amigos	2,34	0,59	0,41	82	662	526

Tabla 2. Índices descriptivos de los ítems

La unidimensionalidad de la escala se evalúa por un análisis de componentes principales en el que el primer componente con un valor propio de 2,05 explica el 41,06% de la varianza total. El segundo componente extraído presenta un autovalor de 0,85 lo que supone una varianza asociada a este componente del 17%. La razón entre la diferencia entre los dos primeros valores propios (2,05-0,85) a la diferencia entre los valores propios segundo y tercero (0,85-0,81) propuesta por Lord (1980) como indicador de unidimensionalidad arroja un valor de 30. Este indicador junto con el porcentaje de varianza asociado al primer componente, que supera los porcentajes utilizados como puntos de corte tanto por Reckase (197) como por Carmines y Zeller (1979) inducen a aceptar la hipótesis de unidimensionalidad de los datos (Elosua y Lopez, 2002).

3.4.2.- Monotonía.

Distribución de las respuestas observadas por grupos de habilidad. La figura 3 muestra la evaluación de la condición de monotonía de las funciones de respuesta entre categorías una vez dividido el continuo de habilidad en cuatro intervalos. Puede comprobarse que todas las funciones de respuesta son monótonas no decrecientes en todos los niveles de habilidad. La base numérica de estas funciones (que no reproducimos por problemas de espacio) concuerda con los requisitos de la monotonía; no existen violaciones para ninguna de las funciones de respuesta.

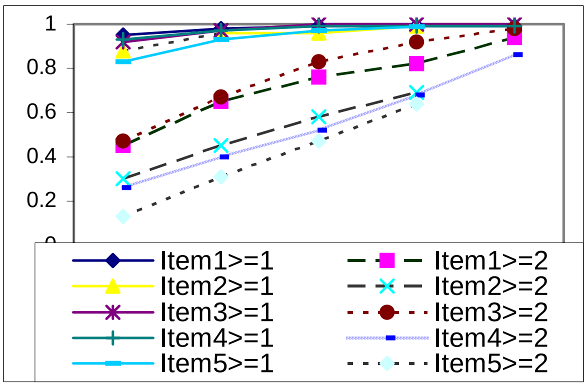

Figura 3. Monotonía y Doble monotonía

Coeficiente de escalabilidad. La tabla 3 muestra los coeficientes de escalabilidad obtenidos para cada ítem (valores de la diagonal) y para cada par de ítems. Todos los valores están próximos al punto crítico de 0,30; los valores mínimos corresponde al par item4-item1

(H_{41}=0,24), item5-item2 (H_{52}=0,23) e item2-item1 (H_{21}=0,28). A modo de ejemplo mostramos los cálculos efectuados para estimar este último valor.

Items	1	2	3	4	5
1	0,31				
2	0,28	0,28			
3	0,33	0,35	0,40		
4	0,24	0,29	0,39	0,34	
5	0,37	0,23	0,55	0,43	0,37

Tabla 3. Coeficientes de escalabilidad

Para computar el coeficiente de escalabilidad se construye para cada par de ítems una tabla (tabla 4) que recoge tanto las frecuencias bivariadas observadas como las frecuencias bivariadas teóricas que se corresponderían con el modelo de independencia para las tres opciones de respuesta a lo ítems. Estas opciones aparecen recodificadas como 0, 1 y 2. Según el patrón de respuestas observado en la tabla, el total de los pasos entre ítems (2 ítems por lo tanto 4 pasos) podrían ordenarse según la secuencia: item1_1 ≥1247, item2_1 ≥ 1206, item1_2 ≥ 951 e ítem 2_2 ≥ 649. Este ordenamiento de los pasos entre categorías da lugar a un ordenamiento teórico de las respuestas. De este orden se derivaría que el patrón teórico 12 correspondiente a las opciones de respuesta de dos ítems, incurriría en un error de acuerdo con los criterios de una escala acumulativa como la escala de Guttman; dado que seleccionar la opción 2 del ítem 2 exigiría haber seleccionado la opción 2 del ítem 1.

El cálculo del coeficiente de escalabilidad (ecuación 9) requiere el número de sujetos en cada uno de los patrones de respuesta bivariados y su ponderación en función de los errores en cada uno de ellos (tabla 4). El coeficiente para estos dos ítems es 0,28.

	N	Frec bivariadas de las categorías de respuesta a dos ítems								
		00	01	02	10	11	12	20	21	22
$X_1 \geq 1$	1247	-	-	-	+	+	+	+	+	+
$X_2 \geq 1$	1206	-	+	-	-	+	+	-	+	+
$X_1 \geq 2$	951	-	-	-	-	-	-	+	+	+
$X_2 \geq 2$	649	-	-	+	-	-	+	-	-	+
Frec. observada.		3	10	9	28	163	105	32	384	534
Frec. esperada		1,1	9,7	11,3	14,7	129,9	151,4	47,1	417,4	486,4
Errores		0	1	3	0	0	1	1	0	0

Tabla 4. Frecuencias observadas, frecuencias teóricas y errores en cada patrón bivariado de respuesta

3.4.3.- Doble monotonía

Una vez ordenados los pasos entre opciones en función de las frecuencias observadas se construye la matriz (P++) (tabla 5) que recoge las frecuencias bivariadas de respuestas para los pasos entre categorías para todos los ítems. En esta tabla es posible comprobar que el orden establecido es monótono no-decreciente en todas las filas y columnas de la matriz. La información que se desprende de la representación gráfica de las funciones de respuesta entre categorías que muestra la figura 3 concuerda con esta resultado.

		P	Item 5	Item 2	Item 4	Item 1	Item 3	Item 5	Item 2	Item 4	Item 1	Item 3
			≥2 0,41	≥2 0,51	≥2 0,59	≥2 0,75	≥2 0,80	≥1 0,94	≥1 0,95	≥1 0,98	≥1 0,98	≥1 0,98
Item 5	≥2	0,41	0,22									
Item2	≥2	0,51	0,26	0,32								
Item4	≥2	0,59	0,32	0,36	0,41							
Item1	≥2	0,75	0,36	0,42	0,49	0,61						
Item3	≥2	0,80	0,39	0,45	0,52	0,65	0,68					
Item5	≥1	0,94	0,40	0,49	0,58	0,72	0,77	0,89				
Item2	≥1	0,95	0,40	0,50	0,58	0,72	0,77	0,89	0,91			
Item4	≥1	0,98	0,41	0,50	0,59	0,74	0,79	0,92	0,93	0,96		
Item1	≥1	0,98	0,41	0,50	0,59	0,74	0,79	0,92	0,94	0,96	0,97	
Item3	≥1	0,98	0,41	0,51	0,59	0,74	0,79	0,92	0,94	0,96	0,97	0,97

Tabla 5. Matriz P(++) de frecuencias observadas

3.4.4.- Doble Monotonía fuerte

El ordenamiento de los ítems se comprueba considerando la concordancia entre el orden de las medias aritméticas de cada uno de los ítems recogidas en la tabla 2 y el ordenamiento que queda reflejado en la matriz P++. La secuencia de los ítems es 5-2-4-1-3.

3.4.5 Orden invariante de los ítems en función del sexo

En el estudio de la invarianza en el orden de los ítems relacionada con el género, se evalúan las matrices de proporciones bivariadas (P++) obtenidas de modo independiente en función del sexo. Como resultado concluimos la equivalencia en el ordenamiento entre los pasos de los ítems entre chicos y chicas. Las proporciones de estudiantes del genero masculino que superan cada paso son respectivamente $Item5_2 \geq 43\%$, $Item2_2 \geq 54\%$, $Item4_2 \geq 63\%$, $Item1_2 \geq 78\%$, $Item3_2 \geq 82\%$, $Item5_1 \geq$ 93%, $Item2_1 \geq 96\%$, $Item4_1 \geq 98\%$, $Item1_1 \geq$ 99% e $Item3_1 \geq$ 99%. Las chicas reproducen este patrón en las siguientes proporciones $Item5_2 \geq 40\%$, $Item2_2 \geq 48\%$, $Item4_2 \geq 56\%$, $Item1_2 \geq 72\%$, $Item3_2 \geq 78\%$, $Item5_1 \geq 94\%$, $Item2_1 \geq 97\%$, $Item4_1 \geq 98\%$, $Item1_1 \geq$ 98%. Desde esta perspectiva del estudio del funcionamiento diferencial de los ítems deduciríamos la independencia entre el orden de los ítems en la escala subyacente de autoconcepto social y el género. Lo cual no implica que pudieran existir diferencias de genero en el constructo medido; la diferencia entre las medias aritméticas (MA_{chicos}=13,05; DT_{chicos}.=1,61; MA_{chicas}=12,73; DT_{chicas}=1,79) obtenidas en los dos grupos 0,32. La significatividad de esta diferencia se evalua con la U de Mann-Whitney que arroja un valor de 182219 (p=0,003). Sin embargo si analizamos la magnitud del efecto que la variable sexo ejerce sobre la variable medida con R^2 podemos comprobar que el tamaño del efecto es 0,007. Este resultado es acorde con trabajos sobre el autoconcepto social que refieren la no-diferenciación en esta dimensión del autoconcepto entre chicos y chicas en contraposición a otros aspectos del autoconcepto como el físico o el académico (Gabelko, 1997; Elosua, 2004; Elosua y López, en prensa; Wilgenbusch y Merrel, 1999).

En definitiva la escala analizada cumple con los requisitos del modelo fuerte de doble monotonía. El cumplimiento de la monotonía permite ordenar a los sujetos sobre la variable manifiesta puntuación total obtenida por la suma de los ítems individuales. El cumplimiento de la doble monotonía fuerte permite además un ordenamiento de los ítems sobre el continuo latente autoconcepto social; además la comprobación del ordenamiento invariante de los ítems nos permite concluir la independencia entre el ordenamiento de los ítems y la variable sexo; no existe funcionamiento diferencial. La escala analizada es una escala unidimensional y acumulativa en la que es posible ordenar los ítems en el continuo de acuerdo a el orden de los ítems en el continuo latente permite

4.- Discusión

La medición psicológica puede clasificarse en función de las características métricas de las escalas obtenidas en medición fundamental o representacional y medición de índice. La primera se caracteriza por establecer una correspondencia biunívoca entre la propiedad medida y la escala de medición; en la segunda esta relación es sólo unívoca (Dawes, 1972). En este hecho reside una de las diferencias principales entre ambos acercamientos, es decir, en la información contenida en la escala métrica referida a la escala psicológica. ¿Qué tipo de inferencias son posibles sobre el objeto medido a partir del conocimiento de su valor escalar? ¿Cuál es el conocimiento sobre la actitud de un sujeto que incorpora una determinada puntuación empírica obtenida en una escala Likert? La asunción mínima de correspondencia de orden entre valores empíricos y valores subyacentes no siempre se cumple. Por ejemplo ante dos ítems politómicos (*i* y *j*) con escalas de respuesta 1, 2, 3 y 4 las opciones de dos sujetos podrían ser 2-3, y 3-2. En esta situación la conclusión sería que ambos sujetos presentan el mismo valor en el rasgo, pero hasta que punto es plausible esta inferencia? Si analizamos individualmente cada uno de los ítems, podrían ser diferentes las conclusiones?

Las inferencias a partir de los valores escalares dependen del grado de cumplimiento de las propiedades de orden estocástico, y éstas están asociadas a diferentes propiedades de los modelos psicométricos de respuesta al ítem. Su conocimiento resulta básico tanto para seleccionar un determinado modelo en función de los requerimientos que se plantee el investigador, como para situar las inferencias derivadas de las puntuaciones obtenidas en un marco teórico y práctico correcto.

Entre el total de modelos disponibles el modelo fuerte de doble monotonía presentado en este trabajo cubre un espacio entre el modelo clásico de tests y los modelos paramétricos de respuesta al ítem que resultan más restrictivos. Permite obtener información métrica de orden referente tanto a los sujetos como al ordenamiento de los ítems en condiciones en las que se disponen de pocos ítems y en situaciones en que los datos no se ajustan a modelos paramétricos. Por ejemplo el ajuste de los datos presentados en este trabajo al modelo de respuesta graduada de Samejima (1997) fue malo. La estimación de los parámetros por el método de máxima verosimilitud marginal para cada uno de los ítems fue: a_1=1,23; b_{11}=-3,84; b_{12}=-1,14; a_2=0,91; b_{21}=-3,64; b_{22}=-0,06; a_1=2,17; b_{31}=-2,73; b_{32}=-1,08; a_4=1,35; b_{41}=-3,29; b_{42}=-0,39; a_5=1,60; b_{51}=-2,27; b_{52}=0,31. El ajuste al modelo se evaluó con el estadístico Chi-cuadrado que fue estimado con la ayuda de MODFIT. Todos los valores resultaros significativos (χ^2_1=143,06; χ^2_2=190,79 ; χ^2_3=33,78; χ^2_4=108,01; χ^2_5=62,192).

Este modelo permite además la evaluación del orden invariante de los ítems. Su cumplimiento es un requisito elemental con vistas a una interpretación correcta de los resultados obtenidos por un test. El hecho de que no puedan ordenarse los ítems porque sus funciones de respuesta se cruzan crearía subordenamientos de ítems diferentes a lo largo de la escala (Figura 2a); es decir, el orden entre ítems dependería del nivel del sujeto en el rasgo. Piénsese en dos ítems cuyas funciones de respuesta se crucen; antes y después de la intersección el orden de los ítems se invertiría, lo cual desde un punto de vista métrico haría depender las características del instrumento de medida de las características del instrumento medido.

Por otro lado, el estudio de la propiedad de orden invariante de los ítems es una herramienta de análisis que se perfila como válida para el estudio del funcionamiento diferencial de los ítems en subgrupos de interés.

Uno de los problemas que presenta el modelo politómico fuerte de doble monotonía es la falta de cumplimiento formal de OEL. Dado que OEL permite realizar inferencias sobre la variable latente, la convierte en una cualidad más interesante que OEM porque permite ordenar a las personas sobre el constructo (no-observable) medido. Esta carencia se traduce en que con este modelo no sería lícito hacer uso de las propiedades métricas de la escala theta (Hambleton y Swaminathan, 1985) y limitaría nuestro conocimiento a propiedades de orden sobre la escala manifiesta (X+). En condiciones prácticas, sin embargo, este hecho no originaría problemas mayores porque habitualmente los resultados se comunican con referencia al orden y no a la distancia, más difícil de interpretar desde una perspectiva psicológica, además los trabajos llevados a cabo hasta el momento coinciden en considerar que las violaciones de esta propiedad métrica se dan en los extremos de la escala X+, lo cual no afectaría al ordenamiento de la mayoría de las personas (van der Ark, en prensa). No disponemos sin embargo de procedimientos empíricos para la evaluación y valoración de esta interesante propiedad que se está convirtiendo en un tema de investigación preferencial.

En definitiva el modelo fuerte de doble monotonía aporta todas las ventajas de los modelos no-parámetricos: proporciona un marco de trabajo más manejable que el que puedan procurar los modelos paramétricos, ofrece herramientas de análisis y evaluación de sus asunciones fáciles de aplicar, y aporta información relevante sobre los datos analizados que puede ser aprovechada tanto en las fases tempranas de construcción de instrumentos de medida donde se dispone de pocos sujetos como en situaciones donde los modelos paramétricos presentan problemas de ajuste.

5.- Referencias

Andrich, D. (1978) A rating formulation for ordered response categories. *Psychometrika,* 43, 561-573.

Carmines, E. G.y Zeller, R. A. (1979). *Reliability and validity assessment*. Londres: Sage.

Dawes, R.M. (1972) *Fundamentals of attitude measurement*. New York: John Wiley & Sons.

Elosua, P. (2004) AFA-A autokontzeptu eskalaren euskaratzea. *Tantak, 32, 39-56.*

Elosua, P.y López, A. (2002). Indicadores de dimensionalidad para ítems binarios. *Metodología de las Ciencias del Comportamiento* **4**(1), 121-137.

Elosua, P.y López, A. (en prensa). Adaptation to basque of a scale about self-concept. Validity Evidences.

Gabelko, N.H. (1997) Age and gender differences in global, academic, social and atletic self-concepts in academicallu talented students. *Paper presented at the Annual Meeting of the American Educational Research Association*: Chicago.

Glas, C., J. Scheerens, et al. (2003). *Educational Evaluation and Monitoring: A systematic approach*, Swets and Zeitlinger.

Grayson, D.A. (1988) Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika,* 53, 383-392.

Guttman, L. (1950) The basis of scalogram analysis. En S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarssfeld, S.A. Star, y J.A. Clausen (Eds.) *Measurement and prediction* (pag. 60-90). Princeton, NJ: Princeton University Press.

Hambleton, R. y Swaminathan, H. (1985) *Item Response Theory. Principles and applications.* Boston: Kluwer Nijhoff Publishing.

Hemker, B.T., Sijtsma, K., Molenaar, I.W. y Junker, B.W. (1996) Polytomous IRT models and monotone likelihood ratios of the total score. *Psychometrika*, 61, 679-693.

Junker, B.W. y Sijtsma, K. (2001). Nonparametric Item Response Theory in Action: An overview of the Special Issue. *Applied Psychological Measurement*, 25(3), 211-220.

Loevinger, J. (1947), A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs,* 61, 4.

Loevinger, J. (1948) The technique of homogenous test compared with some aspects of “scale analysis” and factor analysis. *Psychological Bulletin*, 45, 507-530.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA.

Lord, F.M. y Novick, M.R. (1968) *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Masters, G. (1982) A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Molenaar, I.W: (1991). A weighted Loevinger H-coefficient extending Mokken scaling to multcategory items. *Kwantitatieve Methoden*, 37, 97-117.

Molenaar, I.W. (1997) Nonparametric Models for Polytomous Responses. En van der Linden, W.J. y Hambleton, R. (Eds.), *Handbook of Modern Item Response Theory* (pag. 369-380) New York: Springer.

Molenaar, .W. y Sijtsma, K. (2000) MSP5 for Windows. Groningen: ProGAMMA.

Mokken, R.J. (1971) *A theory and Procedure of Scale Analysis: With Application in Political Research*. New York: Walter de Gruyter.

Mokken, R.J. (1997) Nonparametric Models for Dichotomous Responses. En W.J. van der Linden . y R.K. Hambleton, R. (Eds.), *Handbook of Modern Item Response Theory* (pag.351-368) New York: Springer.

Mokken, R.J. y Lewis, C. (1982) A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 427-430.

Moolenar, W. y Sijtsma, K. (2000) *MSP5 for windows.* Groningen: iecProGAMMA.

Musitu, G., García,, F. y Gutiérrez, M. (1997) *AFA. Autoconcepto Forma-A.* Madrid: TEA.

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

Rosenbaum, P.R. (1984) Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika*, 49, 425-435.

Samejima, F. (1969) Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika Monograph*, 17, 1-100.

Samejima, F. (1997). Graded Response Model. En W. J. van der Linden and R. K. Hambleton. (Eds.) *Handbook of modern item response theory*. (pag. 85-100).New York: Springer-Verlag.

Sijtsma, K. y Hemker, B.T. (1998) Nonparametric Polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika*, 63(2) 183-200.

Sijtsma, K. y Moolenar, W. (2002) *Introduction to nonparametric item response theory*. Thousand Oaks: Sage.

Thurstone, L.L. (1927) A law of comparative judgment *Psychological Review*, 34, 273-286.

van der Ark, L.A. (2001). Relations and Properties of Polytomous Item Response Theory Models. *Applied Psychological Measurement*, 25(3), 273-282.

van der Ark, L.A. (en prensa). Stochastic ordering by the latent trait by the sum score under various polytomous IRT models. *Psychometrika.*

van Schuur, W.H. (2003) Mokken scaly analysis: between the Guttman scale and parametric item response theory. *Political Analysis*, 11, 139-163.

Wilgenbusch, T. y Merrell, K.W. (1999) Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quaterly*, 14(2), 101-120.