

## APLICACIÓN DEL MODELO DE ESCALAS DE CLASIFICACIÓN PARA LA MEDICIÓN DE LA “ACCIÓN POLÍTICA NO CONVENCIONAL”

Antonio J. Rojas Tejada

*Instituto Psicosociológico Andaluz de Investigaciones*

### RESUMEN

La utilización de escalas para medir variables psicosociales es muy habitual en las investigaciones mediante encuestas. Comprobar la escalabilidad de los datos es muy importante para una correcta interpretación de los resultados procedentes de encuestas. Cuando usamos una escala en una investigación mediante encuestas, se debería mostrar la escalabilidad de los datos ya que a partir de ello podría conocerse la dimensionalidad del continuo propuesto y el nivel de escala de medida conseguido. Este trabajo presenta la aplicación del Modelo de Escalas de Clasificación para evaluar la escalabilidad de los datos procedentes de encuestas. El Modelo de Escalas de Clasificación se utiliza para estimar la actitud de las personas y la posición de los items en el constructo medido. En este caso, los datos proceden de la escala de “Acción Política No Convencional”, que está incluida en la Encuesta Mundial de Valores de 1955-96 (1.212 sujetos españoles). Los resultados muestran la posibilidad de localizar a personas e items en función de sus valores de escala en un continuo psicosocial. Así, el Modelo de Escalas de Clasificación se muestra como una forma viable y factible de abordar el escalamiento de variables psicosociales con items politómicos.

*Palabras clave:* Escalamiento, Escalas de Clasificación, Items politómicos, Modelo de Escalas de Clasificación, Acción Política No Convencional.

La información sobre los resultados de una encuesta suele ir acompañada de una serie de datos que intentan mostrar su calidad. En la mayoría de los casos, esta información se refiere casi exclusivamente al error muestral, tamaño de la muestra y procedimiento de muestreo (González, Padilla y Pérez, 1998). Pero esta información es a todas luces insuficiente para garantizar la calidad de los resultados de una encuesta (Groves, 1989).

En las encuestas existe una utilización generalizada de distintas escalas con la finalidad de obtener medidas de diferentes variables psicosociales (p.e. actitudes, creencias, ideologías, etc.), pero es práctica mucho menos generalizada comprobar la escalabilidad de los datos en el continuo psicosocial que se pretende medir.

El hecho de comprobar la escalabilidad de los datos, es decir, comprobar el grado en que los ítems y/o sujetos pueden ser localizados en un continuo psicosocial, es uno de los aspectos más importantes de la medición, ya que a partir del estudio de la escalabilidad podría examinarse la dimensionalidad de los continuos planteados y el nivel de medida de las escalas resultantes. De esta forma, se avanzaría en la comprensión y, posiblemente, en la explicación de la variable psicosocial sometida a examen, estableciendo para cada respuesta diferenciada por parte de los sujetos, una puntuación en el continuo en cuestión, con considerables garantías métricas (Pérez, Padilla, González y Rojas, 1996). Garantías métricas imprescindibles para garantizar la calidad de las medidas procedentes de las encuestas.

Podemos utilizar múltiples procedimientos para comprobar la escalabilidad de los datos: métodos centrados en los estímulos, en los sujetos y en las respuestas (ítems y sujetos a la vez). En el contexto del escalamiento de respuestas, quizás los Modelos basados en la Teoría de Rasgo Latente son los que mayor relevancia están teniendo en el campo de la medición de variables psicosociales.

El objetivo que persigue el presente trabajo es servirnos de uno de los Modelos basados en la Teoría de Rasgo Latente, como lo es el Modelo de Escalas de Clasificación (MEC), para comprobar la posibilidad de escalar conjuntamente ítems y sujetos en la variable "Acción Política No Convencional"; escala que aparece en la Encuesta Mundial de Valores de 1995-1996 (ASEP, 1999). A partir de la aplicación del Modelo de Escalas de Clasificación al conjunto de datos proporcionados por dicha escala podremos concluir sobre la posibilidad de establecer una medida para el continuo psicosocial "Acción Política No Convencional" donde la medida de los sujetos y de los ítems puedan expresarse en la misma métrica de dicho continuo (con nivel de medida de intervalos) y donde podamos realizar comparaciones entre los ítems, entre sujetos y entre sujetos e ítems de forma invariante.

### **Acción Política No Convencional**

Si bien el concepto de Acción Política debe entenderse dentro del contexto de una sociedad determinada, podemos definir dicha participación como cualquier tipo de

acción realizada por una persona o grupo de personas con el fin de incidir de una u otra manera en los asuntos públicos. Hasta los años 70, los trabajos sobre participación política se centraban fundamentalmente en el comportamiento de voto y en las actividades relacionadas con ello (Clemente y Molero, 1994). Sin embargo, desde el surgimiento de los Nuevos Movimientos Sociales (NMS), las actividades o acciones políticas de las personas han mostrado un cambio.

Como decimos, desde comienzos de los años 60 varias oleadas distintas y consecutivas de NMS (p.e. feminista, ecologista, pacifista, etc.) han modificado la vida social y política en las democracias occidentales (p.e. Brand, 1992, Inglehart, 1991). Y es precisamente las nuevas formas de acción política en donde han destacado estos NMS. Los NMS prefieren influir en las decisiones políticas mediante presiones y mediante el peso de la opinión pública, en lugar de comprometerse directamente con la actividad política convencional (Dalton, Kuechler y Bürklin, 1992). Estos movimientos parecen propensos a llevar a cabo acciones políticas en forma de actividades de protesta, con las que pretenden movilizar la opinión pública y atraer su atención con métodos legales (la mayoría de las veces, aunque no siempre); acciones políticas que han sido denominadas por Offe (1992) como “no convencionales”. Por estos modos no convencionales de participación se entendería, siguiendo a Kaase y Marsh, *“el comportamiento que no se corresponde con las normas de la ley y de costumbre que regulan la participación política bajo un régimen particular”* (cit. en Casquete, 1998). Estas acciones pueden ir desde manifestaciones, boicots, huelgas extraoficiales, bloqueos de tráfico, ocupación de edificios, etc. (Inglehart, 1991). Las formas convencionales de participación son aquellas que sí se ajustan a las normas de la ley y costumbres, tales como participar en campañas o ejercer el derecho al voto (p.e. Morán y Benedicto, 1995; Casquete, 1998).

Tanto interés han despertado estos NMS y su forma de expresión que en la Encuesta Mundial de Valores de 1995-96 (ASEP, 1999) se ha abordado la medida de esta “APNC”. En esta encuesta mundial se incluye una escala de 5 ítems con formato de escalas de clasificación de 3 puntos.

### **Escalas de Clasificación y Modelo de Escalas de Clasificación**

Las escalas de clasificación han sido ampliamente utilizadas en las Ciencias Sociales (p.e. Guilford, 1954; Andrich y Masters, 1988), y esto es así por varias razones, entre las que cabe destacar (Spector, 1992): producen instrumentos de medida con buenas propiedades psicométricas, son relativamente sencillas de desarrollar, son fáciles y rápidas de completar por los sujetos, etc. La característica diferencial de este formato de respuestas es el uso de un conjunto fijo de categorías de respuestas ordenadas como alternativas a un ítem, bien para evaluar actitudes, intereses, intenciones, opiniones, personalidad, etc. Este procedimiento es evidente en la puntuación de ítems de actitud elaborados con el formato tipo Likert (Likert, 1932).

Una de las características más importantes a la hora de puntuar estos ítems es la asignación de números enteros sucesivos a las sucesivas categorías, o dicho de otra forma, este formato requiere una primera identificación de varios niveles ordenados de respuestas en los ítems, y posteriormente, asignar una puntuación parcial a cada una de esas categorías de respuestas. Esta puntuación parcial, que muchas veces aparece junto a las alternativas del ítem, suele oscilar entre 1 y el mayor valor (generalmente comprendido entre los valores 3 y 11), indicando cada valor (en ítems positivos) un incremento en la variable que está midiendo cada ítem, bien sea actitudes, intereses, intenciones, opiniones, personalidad, etc.

En un ítem con formato de escalas de clasificación podemos identificar  $k$  categorías de respuestas (o niveles de ejecución), que vendrían dados por el aumento en intensidad en la variable que supone cada alternativa de respuesta: por ejemplo, en el caso de un ítem de cuatro categorías de respuesta: *muy en desacuerdo* (puntuar 1); *en desacuerdo* (puntuar 2); *de acuerdo* (puntuar 3); y, *muy de acuerdo* (puntuar 4). En este caso podríamos definir tres pasos, entendidos conceptualmente como el grado de intensidad necesaria, en la variable que se mide, para llevar a cabo la transición entre dos categorías de respuestas consecutivas de un ítem, que serían:

- 1º paso, definido por la transición entre las categorías 1 y 2
- 2º paso, entre las categorías 2 y 3; y
- 3º paso, entre las categorías 3 y 4

En general, podemos decir que si un ítem tiene  $m+1$  niveles de ejecución (o categorías de respuestas ordenadas), los pasos existentes vienen determinados por  $m$ . De esta forma podemos aplicar este esquema teórico a cualquier campo de la medición donde los ítems presenten alternativas de respuesta ordenadas, como es el caso de la escala de “APNC”.

El MEC (p.e. Andrich, 1978a; 1978b; 1978c; 1982; 1988a; 1988b; Masters, 1980) es un modelo, de la familia de Rasch (1960), que permite tratar la medición de variables psicosociales mediante ítems con formato de escalas de clasificación. El MEC ha mostrado ser de gran utilidad en este contexto, aunque no hay que olvidar que existe un relativamente amplio abanico de modelos de Teoría de Rasgo Latente que pueden aplicarse en este mismo ámbito (Wilson y Masters, 1993). Entre las ventajas del MEC podemos destacar la posibilidad de estimar la intensidad de los ítems y la medida de los sujetos en la variable que se mide de forma invariante respecto a diferentes muestras, es decir, las estimaciones de las intensidades de los ítems son independientes de la muestra de sujetos utilizada para dichas estimaciones y viceversa (p.e. Glass, 1989). Entre los supuestos del modelo están (p.e. Gustafsson, 1980): la unidimensionalidad (lo que implica que todos los ítems miden la misma variable —i.e. Sijtsma, 1998—), la independencia local de ítems y sujetos (esto es, que la respuesta de un sujeto a cualquier ítem del test no resulta afectada por la respuesta dada a otros ítems, e igual con los ítems) y la discriminación homogénea de los ítems (todos los ítems tienen igual poder de discriminación). Pero en el MEC, al igual

que ocurre en casi todo modelo formal, sólo si existe ajuste de los datos podremos gozar de las ventajas que proporciona dicho modelo.

En el MEC se estima la probabilidad que tiene un sujeto de responder a una determinada categoría de un ítem, deduciéndola a partir de la diferencia estimada entre el nivel del sujeto en la variable que se mide y el nivel de intensidad que tienen los ítems utilizados para medir dicha variable. La función utilizada por el MEC es la logística, de forma tal que la probabilidad de que un sujeto  $n$  puntúe en una determinada categoría de respuesta  $x$  en el ítem  $i$  será (p.e. Andrich, 1978a; 1978b; Masters, 1980):

$$\pi_{nix} = \frac{\exp \sum_{j=1}^x [\beta_n - (\delta_i + \tau_j)]}{\sum_{x=1}^{m+1} \exp \sum_{j=1}^x [\beta_n - (\delta_i + \tau_j)]} \quad x = 1, 2, \dots, m+1 \quad (1)$$

donde  $\beta_n$  es la localización o valor de escala del sujeto  $n$  (su valor de medida en la variable);  $\delta_i$  es la localización o valor de escala del ítem  $i$  (su valor de intensidad en la variable que se mide) y  $\tau_m$ , llamado parámetro de umbral entre categorías, es la localización del paso  $m$  relativo al valor de escala del ítem. Una condición que impone el MEC a los parámetros de umbral de los ítems ( $\tau_m$ ) es que éstos se mantengan constantes a través de todos los ítems que formen la escala, y se asume que dependen solamente de las categorías de respuestas propuestas, como es el caso de los tests donde todos los ítems tienen idénticas categorías de respuestas (p.e. tipo Likert, escalas de clasificación, etc). Esta característica lo diferencia del Modelo de Crédito Parcial, ya que en este último no se asume que los parámetros de umbral sean idénticos en todos los ítems, es decir, cada paso  $j$  del ítem  $i$  vendría definido por  $\delta_{ij}$ , donde todos los ítems del test no tienen porqué tener idénticas categorías de respuestas.

En el Modelo de Escalas de Clasificación podríamos decir que el parámetro del paso  $j$  del ítem  $i$  ( $\delta_{ij}$ ) se define como  $\delta_{ij} = \delta_i + \tau_{mj}$ , donde todos los parámetros  $\tau_m$  se mantienen constantes en todos los ítems. En el modelo que tratamos la única diferencia entre ítems se debe a su diferente localización ( $\delta_i$ ) en el continuo unidimensional de la variable que se mide (Masters, 1988).

Para terminar, en este trabajo se pretende, mediante la aplicación del Modelo de Escalas de Clasificación : 1) establecer una medida para un continuo psicosocial (variable que se mide); 2) que las medidas de los sujetos y de los ítems puedan expresarse en la misma métrica en dicho continuo (escalamiento centrado en las respuestas); y, 3) que esa métrica puede determinarse a partir de los datos observados (matriz de puntuaciones ítems  $\times$  sujetos). Además, una vez establecida esa medida, podremos realizar comparaciones entre ítems y entre sujetos, sin que dichas comparaciones

dependan de las características de la muestra de sujetos o de los ítems utilizados en el proceso de estimación (Baker, 1992).

## Método

### Sujetos

La muestra utilizada para este trabajo está formada por el total de personas que contestaron la Encuesta Mundial de Valores 1995-96 (ASEP, 1998) en el ámbito español: 1.212. Del total de sujetos el 48,1% está compuesto por hombres y el 51,9% son mujeres; la edad media de los sujetos es de 45,17 años (mediana=43 y moda=18), con un rango de 74 (máxima=92 y mínima=18), y desviación típica de 18,34.

Se eliminaron aquellos sujetos que no contestaron (o se consideraron casos perdidos) a alguna de las preguntas de la escala de "APNC", quedando un total de 1023 sujetos, no variándose sustancialmente los datos de la muestra original: 48,8% hombres y 51,2% mujeres; edad media de 45,15 años (mediana=43 y moda=20), con rango de 74 (máxima=92 y mínima=18), y desviación típica de 18,15.

### Materiales

La escala de "APNC" se ha extraído de la Encuesta Mundial de Valores correspondiente a los años 1995-1996 (ASEP, 1998). Esta escala consta de 5 ítems, que se corresponden con la pregunta 44 de dicha encuesta (variables 118, 119, 120, 121 y 122) (p.e. del Pino y Bericat, 1998). Los ítems vienen expresados en una escala de clasificación de 3 puntos, es decir, con tres categorías de respuestas ordenadas ("*he hecho*", "*podría hacer*", "*nunca lo haría*") que se puntúan con 3, 2 y 1 puntos, respectivamente (se ha incluido una versión en el anexo I).

Para llevar a cabo la aplicación del Modelo de Escalas de Clasificación se ha empleado el programa desarrollado por Wright y Linacre (1992b): BIGSTEPS versión 2.29.

## Resultados

### Ajuste

El ajuste de los datos al MEC se comprueba a través de distintos tipos de tests de ajuste. Masters y Wright (1997) los dividen en tres: 1) tests de ajuste de los ítems; 2) tests de ajuste de los sujetos; y finalmente, 3) tests de ajuste global, que señala el grado de ajuste general del conjunto de datos al modelo (ítems y sujetos).

El programa BIGSTEPS nos proporciona la bondad de ajuste para los tres tipos indicados (Wright y Linacre, 1992a). Los tests de ajuste utilizados en el programa se basan en el análisis de residuos. En el análisis de residuos o residuales se estiman los parámetros de los ítems y de los sujetos, y se hacen predicciones acerca de las res-

puestas de los sujetos a los ítems (puntuaciones estimadas o teóricas). Estos resultados pronosticados se comparan con los datos empíricos. Los residuos son el resultado de la diferencia entre las puntuaciones teóricas y las empíricas. Concretamente, los estadísticos propuestos por Wright y Masters (1982) para el estudio del ajuste de los datos al MEC se basan en los Residuales Cuadráticos Medios No Ponderados —RCMnp— y Residuales Cuadráticos Medios Ponderados —RCMp—. Cuando los datos observados coinciden plenamente con los propuestos por el modelo estos estadísticos tienen valores cercanos a 1. Así valores alejados de 1, muestran que existe desacuerdo entre lo esperado y lo observado. Lunz, Wright y Linacre (1990) han sugerido que la región para considerar un ajuste aceptable oscilaría entre los valores de 0,6 y 1,5 para estos estadísticos (*cit. en Engelhart y Stone, 1998*). Pero estos residuales muestran el inconveniente de no tener en cuenta que la estimación no se realiza con la misma precisión en todos los puntos del continuo, y por ello, es mucho más acertado tomar estos residuales de forma estandarizada ya que de esta manera sí son sensibles a ello. Así, tenemos que cuando estandarizamos el RCMnp pasa a ser el Residual Cuadrático Medio Estandarizado de ajuste externo (“outfit”) y el RCMp pasa a denominarse Residual Cuadrático Medio Estandarizado de ajuste interno (“infit”). Estos nuevos estadísticos se distribuyen según una normal estandarizada, siempre y cuando no existan discrepancias entre los datos observados y los esperados por el modelo.

Por tanto, para considerar si los datos (ya sean de ítems o de sujetos) se ajustan al modelo se debe determinar y estudiar: a) los valores de los RCMnp y RCMp, tanto de ítems como de sujetos, para determinar si están comprendidos entre 0,6 y 1,5, con medias cercanas a 1, tal y como cabe esperar cuando existe buen ajuste; y, b) los valores de “outfit” e “infit” tanto de ítems como de sujetos, para determinar sus medias y desviaciones típicas, ya que cuando existe un buen ajuste de los datos al modelo sus medias deben valer 0 y sus desviaciones típicas ser iguales a 1.

Tabla 1: *Ajuste de los ítems*

ITEM	RCMp	“INFIT”	RCMnp	“OUTFIT”	C.I.T.
1 “Firmar una petición”	1,07	1,35	1,17	2,34	0,28
2 “Secundar boicots”	0,93	-1,22	0,89	-1,29	0,49
3 “Participar en manif.”	0,95	-0,94	0,92	-1,28	0,37
4 “Participar en huelg.”	1,00	0,07	0,87	-1,12	0,51
5 “Ocupar edificios...”	1,10	1,32	1,11	0,68	0,38

Los resultados del ajuste de los ítems, mostrados en la tabla 1, indican valores de RCMnp y RCMp dentro de los límites para considerarlos como indicadores de un buen ajuste. Igual ocurre al fijarnos en los valores de “infit” y “outfit”, ya que para todos los ítems oscilan entre -2 y +2 (salvo el ítem 1 que tiene un valor para “outfit”

ligeramente por encima de +2). En la columna C.I.T. se muestran las correlaciones entre la puntuación en el ítem y la puntuación total en la escala. Se observa que todas son positivas y oscilan entre el valor mínimo de 0,28 y el máximo de 0,51.

Los resultados del ajuste de los sujetos muestran valores de RCMnp y RCMp fuera de los límites para 334 (32,36%) y 381 (37,24%) sujetos de la muestra, respectivamente (ver tabla 2). Al tomar estos estadísticos de forma estandarizada, valores "infit" y "outfit", no ocurre igual, ya que tan sólo 21 (2,05%) y 37 (3,62%) sujetos muestran valores no permitidos.

Tabla 2: *Ajuste de los sujetos*

Valores	RCMp	RCMnp	Valores	"INFIT"	"OUTFIT"
≤0,6	193	295	≤-2	0	0
>0,6 y <1,5	689	642	>-2 y <+2	1002	986
≥ 1,5	141	86	≥ +2	21	37

El ajuste global, basándonos en las medias y desviaciones típicas de los estadísticos de ajuste, nos muestra unos valores de RCMnp y RCMp, tanto para ítems como para sujetos, cercanos a 1 (ver tabla 3). Los valores de "infit" y "outfit" también tienen medias cercanas a 0 y desviaciones típicas cercanas a 1, tal y como cabe esperar cuando existe buen ajuste.

Tabla 3: *Ajuste global (ítems y sujetos)*

		RCMp	"INFIT"	RCMnp	"OUTFIT"
ITEMS	Medias	1,01	0,1	0,99	-0,1
	D.Típicas	-----	1,1	-----	1,4
SUJETOS	Medias	0,99	-1	0,98	0
	D.Típicas	-----	1,1	-----	1

Por tanto, y a modo de conclusión, podemos decir que existe un ajuste más que aceptable del conjunto de datos (ítems y sujetos) al MEC, tal y como indican los resultados mostrados.

### Calibración de los ítems

Debido a que los ítems de la escala de "APNC" tienen todos tres categorías de respuestas, se han estimado dos parámetros para los dos pasos de cada ítem ( $\delta_{1j}$  y  $\delta_{2j}$ ). Recordemos que los parámetros de los pasos  $j$  de los ítems  $i$  ( $\delta_{ij}$ ) se definieron como la intersección entre dos categorías adyacentes, y en el caso del MEC vienen defini-



dos por  $\delta_i$ , que es el valor de escala del ítem  $i$  (su valor de intensidad en la variable que se mide), y por  $\tau_m$ , que es el parámetro de umbral entre categorías, o localización del paso  $m$  relativo al valor de escala del ítem ( $\delta_{ij} = \delta_i + \tau_m$ ).

Los valores de los parámetros de los pasos se recogen en la tabla 4. En ella aparecen los valores de  $\delta_i$  y de  $\tau_m$  de los 5 ítems<sup>1</sup>. Para  $\delta_i$  se puede observar que el rango de valores oscila entre +1,79 para el ítem 5 y -1,84 para el ítem 1. Además, en los Modelos basados en la Teoría de Rasgo Latente, como en el caso del Modelo de Escalas de Clasificación, se proporcionan los errores de estimación para cada ítem, proporcionando información específica sobre la calidad de la calibración de los ítems en diferentes puntos del continuo que se establece para la variable que se mide. Los parámetros del primer paso de cada ítem se obtienen sumando el valor de  $\delta_i$  al de  $\tau_1$ ; y el del segundo sumando el valor de  $\delta_i$  al de  $\tau_2$ .

Tabla 4: Calibración de los ítems

ITEMS	$\delta_i$	Error Estimación	Primer Paso $\delta_{i1}$	Segundo Paso $\delta_{i2}$
1 "Firmar una petición"	-1,84	0,07	-3,34	-0,34
2 "Secundar boicots"	0,59	0,08	-0,91	2,09
3 "Participar en manifestac."	-1,69	0,07	-3,19	-0,19
4 "Participar en huelgas il."	1,14	0,08	-0,36	2,64
5 "Ocupar edificios o fábricas"	1,79	0,1	-0,29	3,29
Valores de los parámetros de umbral: $\tau_1 = -1,50$ y $\tau_2 = +1,50$				

El Modelo de Escalas de Clasificación nos permite dibujar las curvas de probabilidad para cada categoría del ítem, siendo éstas idénticas para todos y difiriendo únicamente en la localización en el continuo de la variable "APNC". Así, y a modo de ejemplo, en la figura 1 y 2 mostramos las curvas de probabilidad de los ítems 1 y 5, respectivamente.

Observamos que las curvas de probabilidad de las categorías de respuestas son similares en ambas figuras, tan sólo que en el ítem 1 está centrada en el valor de escala  $\delta_1 = -1,84$ , y en el ítem 5 alrededor de  $\delta_5 = +1,79$ .

Considerando cómo se puntuó cada categoría de respuesta, podemos decir que el ítem que mayor intensidad tiene en la variable que se mide es el 5 (valor de calibra-

---

<sup>1</sup> Para poder utilizar e interpretar los valores tanto de la intensidad de los ítems como de medida de los sujetos, el programa BIGSTEPS realiza una transformación logarítmica de los datos de forma que ambas quedan en una misma escala, la conocida "escala logit", con media 0 y desviación típica 1. En el texto de Wright y Masters (1982) se lleva a cabo una exposición detallada de cómo se realizan estas transformaciones.

ción de +1,79), siguiéndole los ítems 4, 2, 3 y, por último, el que menor intensidad manifiesta es el 1 (valor de calibración de -1,84).

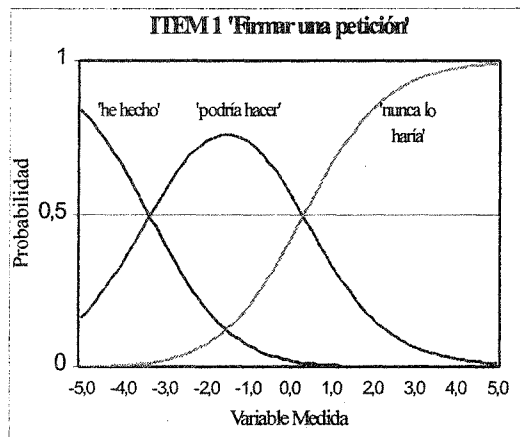


Figura 1: Curva de probabilidad del ítem 1

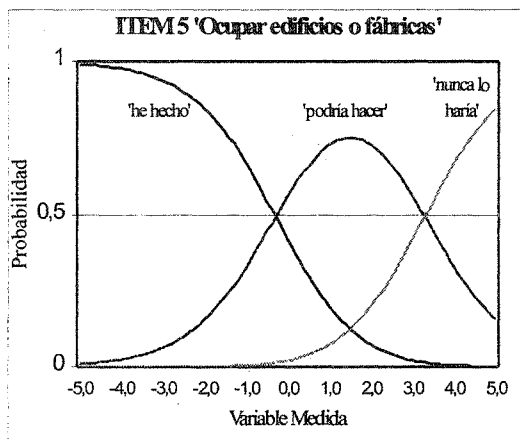


Figura 2: Curva de probabilidad del ítem 5

### Medición de los sujetos

La medida de los sujetos se resume en la tabla 5, donde se muestran los estadísticos de los parámetros correspondientes a sus valores en la variable medida. El promedio de las medidas está en el valor  $-2,53$  (d.t.=2,16). Lo más destacado es comprobar cómo la distribución no es simétrica, habiendo un sesgo a favor de los sujetos que tienen puntuaciones menores a la media en “APNC”.

Tabla 5: *Medidas de los sujetos*

	Media	Mediana	Moda	Desv. Típ.	Rango		Curt.	Asim.
					Mínimo	Máximo		
<b>SUJETOS</b>	-2,53	-2,47	-4,71	2,16	-4,71	4,58	-0,04	0,78

Una limitación del Modelo de Escalas de Clasificación en la medición de los sujetos está en cuanto a la estimación de la medida de los sujetos extremos, es decir aquellos que puntúan alto en todos los ítems (33333) o aquellos que tienen puntuaciones bajas (11111). Para estos sujetos el modelo no puede localizar exactamente el lugar donde situarlos en el continuo, solo sabemos que tienen puntuaciones mínimas o máximas. En nuestro caso, el modelo asigna a estos sujetos el máximo valor, que es de 4,58; y el mínimo valor, que es de  $-4,71$ , respectivamente. Estos sujetos no intervienen en la calibración de los ítems.

### Mapa de ítems y sujetos

Para tener una visión global de cómo se distribuyen los ítems y sujetos en el continuo que mide la escala de “APNC” se puede generar un *mapa de la distribución de sujetos e ítems*. En la figura 3 se presentan los ítems y los sujetos, calibrados y medidos, respectivamente, en la misma escala *logits* (que aparece en el eje de ordenada desde el valor de +4 a -4).

Todas las medidas de los sujetos aparecen a la izquierda de la figura 3. En ella se puede contemplar su distribución a lo largo de la variable medida. Se puede observar también la asimetría indicada anteriormente.

Todos los ítems aparecen en la zona derecha en tres columnas, registrando el rango de actuación de los ítems.

En la columna central (“*ítems media*”), se localizan las calibraciones medias de los ítems; podríamos considerar que esta es la intensidad del ítem tomado como un todo. Pero recordemos que los ítems utilizados en este trabajo tienen tres categorías de respuestas y que se distribuyen a lo largo del continuo en función de los valores de intensidad de cada paso del ítem. Pues bien, en la columna de la izquierda (“*ítems bajo*”) se presenta la distribución de los ítems según el funcionamiento de las catego-

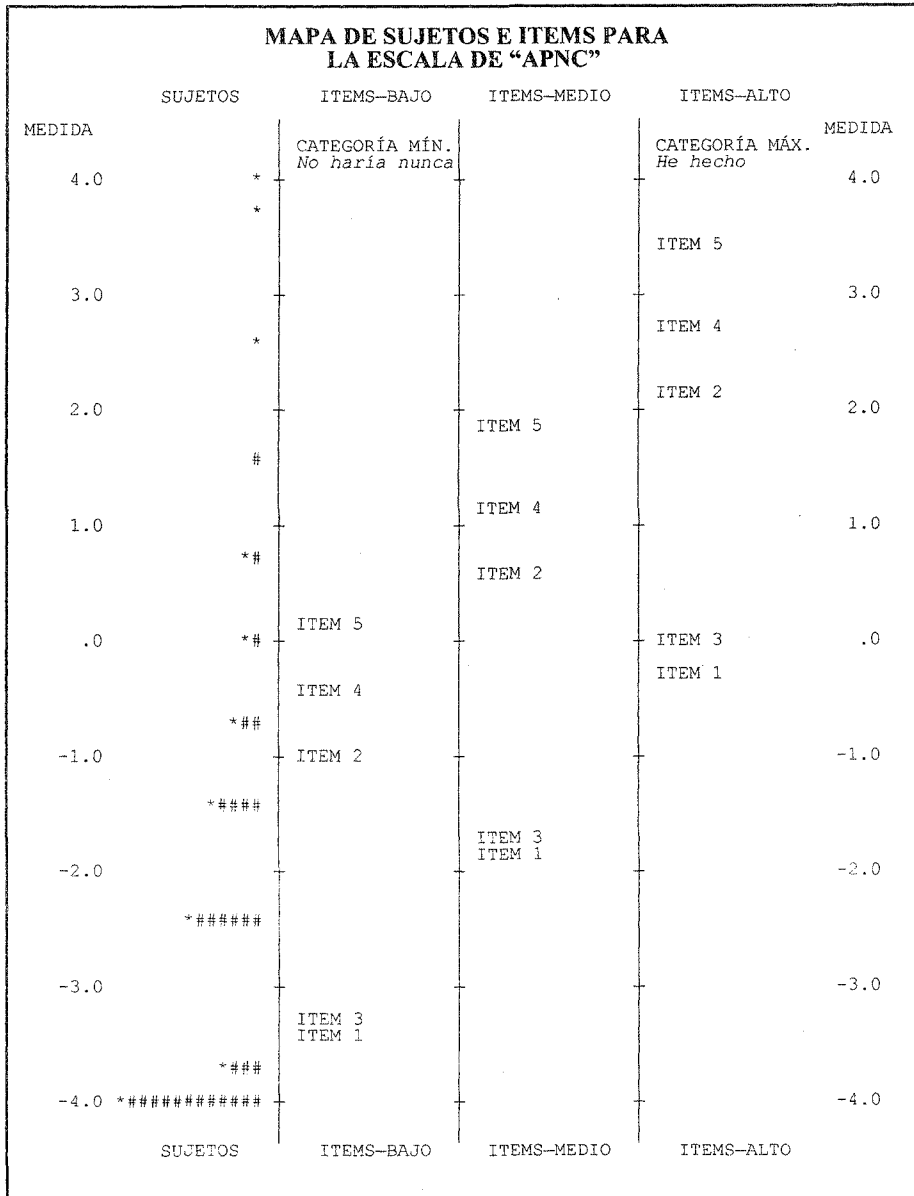


Figura 3: Mapa de sujetos e items

rías de respuestas con menor puntuación (1 punto). Esta columna ofrece una idea de hasta donde actúan en el continuo las categorías de respuestas “No haría nunca” de los diferentes items.

En la columna de la derecha (“*items alto*”) se muestra la distribución de los items según el funcionamiento de las categorías de respuestas con mayor puntuación (3 puntos). Esta columna refleja hasta donde actúan, en el continuo que se mide, las categorías de respuestas “He hecho” de los distintos items.

En definitiva, la conjunción de las tres columnas nos proporciona el rango de actuación de los items (funcionamiento mínimo, máximo y medio) teniendo en cuenta sus categorías de respuestas con mayor y menor puntuación, así como su puntuación media.

Lo importante de este mapa, donde se representan la distribución de las medidas de los sujetos y el rango de medida de los items, es comprobar si este rango de actuación de los items es adecuado. Con el conjunto de items que nos ocupa, pensados para medir “APNC” en la población general, podemos decir que reflejan un rango de funcionamiento aceptable para recoger la variación de los sujetos en esta variable. El rango de funcionamiento de un item lo dan las curvas de probabilidad de las categorías de respuesta que tenga dicho item, pero el lugar que se representa en este mapa es el valor medio de la categoría en la variable que se mide, y no hay que dejar de recordar que estas curvas son asintóticas en los extremos respecto a la variable que se mide (obsérvese figuras 1 y 2).

También podemos escribir en forma de tabla la información, para los items, que se presenta en la figura anterior. De esta manera, podemos reescribir la tabla 4 para mostrar los valores que indican el rango de funcionamiento de todos los items. En la tabla 6 aparecen dichos valores para todos los items, ordenados por el valor de escala del item o funcionamiento medio.

Tabla 6: *Rango de funcionamiento de los items*

ITEMS	Funcionamiento	Funcionamiento Medio	Funcionamiento Máximo
1 “Firmar una petición”	-3,34	-1,84	-0,34
3 “Participar en manif.”	-3,19	-1,69	-0,19
2 “Secundar boicots”	-0,91	0,59	2,09
4 “Participar en huelg.”	-0,36	1,14	2,64
5 “Ocupar edificios...”	-0,29	1,79	3,29

En nuestro caso el rango de funcionamiento iría desde la categoría menor puntuada (“no lo haría nunca”) del item 1 con  $-3.34$ , hasta la categoría mayor puntuada (“he hecho”) del item 5 con  $+3.29$  (en logits). A la vista del mapa, quizás podríamos pensar que esta escala podría mejorarse si incluyéramos algunos items con localiza-

ciones de sus pasos menores a  $-3.34$ , lo que posibilitaría reflejar y discriminar entre las medidas de todos los sujetos que se encuentran por debajo de esa posición.

## Conclusiones

En este trabajo se ha llevado a cabo una aplicación del Modelo de Escalas de Clasificación a los datos procedentes de administrar los ítems elaborados para medir "APNC", incluidos dentro de la Encuesta Mundial de Valores 1995-96, a una muestra de 1212 sujetos, con el fin de poder determinar si es posible situar a ítems y a sujetos en el mismo continuo psicosocial propuesto.

Respecto a los resultados del ajuste de los datos al MEC se ha comprobado que se cumplen de forma adecuada los requisitos exigidos a los estadísticos de ajuste (tanto de los sujetos como de los ítems). En nuestro caso existe evidencia para afirmar que el modelo de medida proporciona una adecuada descripción de los datos, lo que conlleva al menos a tres consideraciones:

1. El conjunto de ítems define una representación unidimensional de los sujetos e ítems en la variable que se mide.
2. La medida de la variable obtenida tiene un nivel de medida de intervalos y, por tanto, se pueden interpretar las distancias entre sujetos, entre ítems, y entre sujetos e ítems.
3. Las localizaciones de los ítems son invariantes respecto a diferentes muestras de sujetos, y las localizaciones de los sujetos en la escala son invariantes respecto al conjunto de ítems usados en la medida de la variable.

Los resultados de la aplicación del Modelo de Escalas de Clasificación, guiados por el programa BIGSTEPS, nos ha proporcionado:

4. Información sobre la *calibración de los ítems*, donde se han determinado los parámetros de los pasos de los ítems así como la intensidad global de cada ítem. También se ha obtenido información sobre la *medida de los sujetos*.
5. Información sobre el funcionamiento de todos los ítems y del conjunto de sujetos, utilizando la misma escala de medida. Los rangos de funcionamiento de los ítems (y de sus categorías de respuestas) han mostrado que reflejan valores de adhesión lo suficientemente amplios como para ser considerados adecuados. Además, se ha representado la distribución de las medidas de los sujetos.

Como conclusión, podemos decir que los ítems y los sujetos tratados pueden ser escalados en una variable psicosocial. En nuestro caso el contenido de dicha variable ha sido la "APNC", cuestión que ha de ser investigada en posteriores estudios de validación.

Debemos acentuar la necesidad de comprobar la escalabilidad de los datos, ya que es un requisito fundamental para que la información que se obtiene de las encuestas sea de calidad. Así, sería recomendable que en toda encuesta, donde se utilicen

expresamente escalas de medida compuestas por diferentes ítems y a partir de las cuales se obtienen las puntuaciones o valores de escala de los sujetos, se incluya información, además de la habitual relativa al error de precisión, nivel de confianza, tasa de respuesta, etc., sobre la calidad de la medida que proporciona la escala. Y no nos estamos refiriendo exclusivamente a informar sobre la fiabilidad de las puntuaciones (cosa que es muy recomendable), sino a la información sobre la escalabilidad de los datos (sujetos e ítems). Además, a partir de este escalamiento podemos obtener los valores de escala de los ítems que componen la escala, cuestión muy útil para el análisis de los mismos. Incluso podría informarse sobre el funcionamiento no ya de los ítems, sino de cómo lo hacen las distintas categorías de respuesta, tal y como ha sido el caso tratado aquí. De hecho, en el presente trabajo hemos querido presentar cómo el MEC es una forma viable y factible de abordar el escalamiento de variables psicosociales con ítems politómicos. Esto nos debe hacer reflexionar sobre la posibilidad de utilizar este modelo frente a formas clásicas de abordar el escalamiento, sobre todo, pensando en las ventajas que supone la aplicación de este modelo (p.e. invarianza de las medidas, escalamiento conjunto de ítems y sujetos, unidimensionalidad, etc.).

No nos cabe duda de que la utilización de los modelos basados en la Teoría de Rasgo Latente, como lo es el MCE, debe ser de gran utilidad para mejorar la forma de abordar la medida de las escalas incluidas en las encuestas, y por consiguiente, de las informaciones que se extraen de ellas.

### Referencias

- Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, Vol.43. 561-573.
- Andrich, D. (1978b). Scaling attitude items constructed and scored in the Likert Tradition. *Educational and Psychological Measurement*, Vol.38. 665-680.
- Andrich, D. (1978c). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, Vol.2, N°4, 581-594.
- Andrich, D. (1982). Using latent trait measurement models to analyse attitudinal data: a synthesis of viewpoints. En D. Spearritt (Ed.). *The improvement of measurement in education and psychology*. Burwood, Victoria: Australian Council for Educational Research. 89-126.
- Andrich, D. (1988a). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement In Education*, 1(4). 363-378.
- Andrich, D. (1988b). *Rasch models for measurement*. Beverly Hills. Sage Publications.

- Andrich y Masters (1988). Ratin scales analysys. En J.P. Keeves (Ed.). *Educational research, methodology and measurement: an international handbook*. Elmsford, N.Y.: Pergamon Press. 297-303.
- ASEP (1999). *Encuesta Mundial de Valores 1995-96*. [CD-ROM]. Madrid: Análisis Sociológico Económico y Político.
- Baker, F.B. (1992). *Item response theory: parameter estimation techniques*. New York: Marcel Dekker, Inc.
- Brand, K-W. (1992). Aspectos cíclicos de los nuevos movimientos sociales: fases de crítica cultural y ciclos de movilización del nuevo radicalismo de clases medias. En R.J. Dalton y M. Kuechler (Comp.). *Los nuevos movimientos sociales*. Valencia: Generalitat Valenciana. p. 45-69.
- Casquete, J. (1998). *Política, cultura y movimientos sociales*. Bilbao: Bakeaz.
- Clemente, M. y Molero, F. (1994). Estudio psicosocial de la delincuencia y participación política no convencional. En J.F. Morales y otros. (Coor.). *Psicología social*. Madrid: MacGraw-Hill. p. 845-873.
- Dalton, R.J., Kuechler, M. y Bürklin, W. (1992). El reto de los nuevos movimientos. En R.J. Dalton y M. Kuechler (Comp.). *Los nuevos movimientos sociales*. Valencia: Generalitat Valenciana. p. 19-42.
- Del Pino, J. y Bericat, E. (1998). *Valores sociales en la cultura andaluza. Encuesta Mundial de Valores. Andalucía 1996*. Madrid: CIS.
- Engelhard, G. y Stone, G. E. (1998). Evaluating the quality of ratings obtained from standard-setting judges. *Educational and Psychological Measurement*, Vol.58, N°2, 179-196.
- González, A., Padilla, J.L. y Pérez, C. (1998). La validez en las investigaciones mediante encuestas. En V. Manzano y M. Sánchez (Comp.). *Investigación del Comportamiento*. Sevilla: IPAL.
- Groves, R.M. (1989). *Survey errors and survey costs*. New York: Wiley.
- Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Gustafsson, J. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33. 205-233.
- Inglehart, R. (1991). *El cambio cultural en las sociedades industriales avanzadas*. Madrid: CIS.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-54.
- Lunz, M.E., Wright, B.D. y Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement In Education*, 3 (4). 331-345.
- Masters, G.N. (1980). A Rasch model for rating scales. *Dissertation Abstracts International*, 41, 215A-216A.
- Masters, G.N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, 1(4). 279-297.



- Masters, G.N. y Wright, B.D. (1997). The partial credit model. En W. J. van der Linden y R.K. Hambleton (Eds.). *Handbook of modern item response theory*. Cap.6. 101-121.
- Morán, M.L. y Benedicto, J. (1995). *La cultura política de los españoles*. Madrid: C.I.S.
- Offe, C. (1996). *Partidos políticos y nuevos movimientos sociales*. Madrid: Editorial Sistema.
- Pérez, C., Padilla, J.L., González, A. y Rojas, A.J. (1996). El uso de las técnicas de escalamiento para la definición y medida del continuo psicosocial individualismo-colectivismo. *Boletín de Psicología*. Nº 52, 7-23.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. (Reprinted by the Chicago University Press, 1980).
- Sijtsma, K. (1998). Methodology review: nonparametric IRT approaches to the analysis of dichotomous items scores. *Applied Psychological Measurement*, Vol.22, Nº1, 3-31.
- Spector, P.E. (1992). *Summated rating scale construction*. Newbury Park, California: SAGE Publications, Inc.
- Wilson, M. y Masters, G.N. (1993). The partial credit model and the null categories. *Psychometrika*, nº 58, 87-99.
- Wright, B.D. y Linacre J.M. (1992a). *A user's guide to BIGSTEPS*. Chicago. Mesa Press.
- Wright, B.D. y Linacre, J.M. (1992b). *BIGSTEPS ver. 2.2. Computer Program*. Chicago: Mesa Press.
- Wright, B.D. y Masters, G.N. (1982). *Rating scale analysis*. Chicago. Mesa Press.

**Anexo I**  
*Escala APNC de 5 ítems*

Voy a leerle algunas de las diferentes formas de acción política que lleva a cabo la gente, y me gustaría que me dijera, para cada una de ellas, si ha hecho Ud. alguna de estas cosas, si podría hacerlas en un futuro o si nunca las haría bajo ninguna circunstancia

	<u>He hecho</u>	<u>Podría hacer</u>	<u>Nunca lo haría</u>	<u>No sabe</u>
1. Firmar una petición	1	2	3	9
2. Secundar boicots	1	2	3	9
3. Participar en manifestaciones legales/autorizadas	1	2	3	9
4. Ocupar edificios o fábricas	1	2	3	9