

## ANÁLISIS DE LAS ALTERNATIVAS DE RESPUESTAS INTERMEDIAS MEDIANTE EL MODELO DE ESCALAS DE CLASIFICACIÓN<sup>1</sup>

Antonio J. Rojas Tejada  
Juan S. Fernández Prados  
*Universidad de Almería*

### RESUMEN

El presente trabajo trata sobre uno de los inconvenientes más importantes en los ítems de escalas de clasificación (p.e. tipo Likert): el problema de la categoría de respuesta intermedia. El objetivo del presente artículo es llevar a cabo una aplicación del “modelo de escalas de clasificación” (MEC) para obtener información sobre el funcionamiento de estas categorías, y por tanto, de su relevancia dentro de los ítems. Para ello, se ha llevado a cabo un estudio donde se ha utilizado la “escala de timidez” de Cheek y Buss, donde se ha dividido al conjunto de sujetos en dos grupos en función de la información que proporcionan dichos sujetos sobre cómo han interpretado la categoría ‘*Ni de acuerdo ni en desacuerdo*’: bien como una categoría que expresa una posición intermedia bien como una categoría que expresa duda o indecisión. Los resultados muestran que en el primer grupo las curvas características de las categorías de respuestas producen una región de respuesta para esta alternativa. Ello no ocurre en el caso del segundo grupo, donde las curvas características de las categorías de respuestas no originan dicha región de respuesta más probable. Como conclusión, podemos decir que en el primer caso la categoría intermedia es relevante y tiene sentido incluirla, mientras que para el segundo grupo la categoría intermedia no es relevante ya que no está proporcionando esa información, y quizás podría eliminarse.

*Palabras clave:* Categorías de Respuestas Intermedias, Escalas de Clasificación, ítems tipo Likert, Modelo de Escalas de Clasificación.

---

<sup>1</sup> Este trabajo forma parte de un proyecto financiado por el CICYT dentro del Programa Nacional de Promoción General del Conocimiento (PB98-1156).

## Introducción

Los ítems que presentan escalas de clasificación son ampliamente utilizados en las Ciencias Sociales. La característica diferencial de este formato de respuestas es el uso de un conjunto de categorías de respuestas, iguales para todos los ítems de la escala. Un ejemplo de este tipo son los ítems elaborados con el formato tipo Likert (Likert, 1932), en el que los sujetos exponen su grado de acuerdo o desacuerdo respecto a lo que dice el enunciado del ítem (Andrich, 1978b; 1988a). Concretamente, el formato tipo Likert se ha convertido en uno de los formatos de ítems de escalas de clasificación más utilizados, tanto en el contexto de la medición psicológica (Drasgow, 1995) como en otros ámbitos: estudios de mercado y opinión, marketing, etc. (Cañadas y Sánchez, 1998).

Las escalas de clasificación son muy utilizadas por varias razones (Spector, 1992): producen escalas con buenas propiedades psicométricas, son relativamente sencillas de desarrollar, son fáciles y rápidas de completar por los sujetos, etc. En las escalas de clasificación se presenta un conjunto fijo de categorías de respuestas ordenadas como alternativas a un ítem para evaluar actitudes, personalidad, opiniones, etc. La tarea de los sujetos implica responder si están bien 'de acuerdo', bien 'en desacuerdo' con lo propuesto por el ítem, así como la magnitud de ese acuerdo o desacuerdo. Habitualmente, las categorías de respuestas son simétricas y se presentan alrededor de una categoría neutral o intermedia.

Una de las características más importantes a la hora de puntuar ítems con categorías de respuestas ordenadas es asignar 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 (siendo este valor desde 3 hasta 11).

Entre los inconvenientes de las escalas de clasificación destacan (Andrich y Masters, 1988): la determinación del número óptimo de categorías que deben tener los ítems y el problema de la categoría intermedia, además de otros problemas no menos importantes como por ejemplo el problema de la familiaridad de las categorías de respuesta (Cañadas y Sánchez, 1998; Durán, Ocaña, Cañadas y Pérez, 2000).

Centrándonos en el segundo de los problemas, podemos decir que la categoría intermedia en un ítem (categoría habitualmente indicada como 'ni de acuerdo ni en desacuerdo', 'ni positivo ni negativo', 'indiferente', etc.) se ha convertido en un 'cajón de sastre' donde se incluyen tanto las respuestas de las personas que tienen dudas o son indecisas respecto a lo que demanda el ítem, como las respuestas de aquellas personas que toman dicha categoría como una categoría neutral o intermedia entre las propuestas, e incluso se puede dar que algunas personas la elijan debido a que no han comprendido bien el ítem (DuBois y Burns, 1975).

A partir de lo dicho cabe plantearse una cuestión ¿cuál es el objetivo que nos lleva a incluir la categoría intermedia o neutral en un ítem? Obviamente se puede incluir con varios propósitos, pero quizás los más importantes son: 1) incluirla para que sea elegida por las personas indecisas o que presenten dudas (donde lo ideal es que sea elegida por pocas personas); 2) incluir esta alternativa para que funcione como una categoría intermedia o neutra, donde se asigne una puntuación intermedia entre la categoría menor y la mayor, y donde está latente el supuesto de que las distintas categorías están ordenadas por intensidad (Spector, 1992). El incluir dicha categoría en los ítems implica que, en las instrucciones de la escala, debe indicarse expresamente la finalidad de su uso. No obstante, hay que comentar expresamente que en las indicaciones la finalidad que cumple esta categoría no garantiza que funcione bien para el objetivo propuesto, y además no es usual explicar el uso que debe hacerse de dicha categoría. En cualquier caso, algunos autores (p.e. Andrich y Masters, 1988) proponen que se incluya esta categoría de respuesta intermedia con el fin de observar cuántas se registran, y eliminarla si resultan ser pocas.

Es justo en este último punto donde el Modelo de Escalas de Clasificación —*Rating Scale Model*— (p.e. Andrich, 1978a; 1978b; 1978c; 1982; 1988a; 1988b; Masters, 1980) se muestra de gran utilidad, aunque existe un relativamente amplio abanico de modelos de Teoría de Rasgo Latente que pueden aplicarse en este contexto (Wilson y Masters, 1993).

El Modelo de Escalas de Clasificación (MEC) es un modelo politómico de la familia de Rasch (1960) que permite tratar la medición de variables psicosociales donde se utilizan ítems con formato de escalas de clasificación. Entre las ventajas de este modelo 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), ventajas que sólo existen si se consigue ajuste de los datos al modelo. Además, al ser politómico, podemos estudiar el funcionamiento de cada categoría de respuesta de los ítems, cuestión que, como veremos más adelante, es fundamental para el objetivo que pretende el presente trabajo.

A los ya habituales supuestos de los modelos basados en la Teoría de Rasgo Latente como son la unidimensionalidad (que implica que todos los ítems miden la misma variable —p.e. Sijtsma, 1998—), y 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) se les une en este caso la discriminación homogénea de los ítems (todos los ítems tienen igual poder de discriminación).

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):

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

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  $i$ , 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 (Wright y Masters, 1982), como es el caso de los tests donde todos los ítems tienen idénticas categorías de respuestas (p.e. tipo Likert, escalas de clasificación). En el MEC podríamos decir que el parámetro del paso  $j$  del ítem  $i$  ( $\delta_{ij}$  o valor de intensidad en la variable que se mide) se define como  $\delta_{ij} = \delta_i + \tau_m$ , 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), es decir, a la intensidad con la que mide cada ítem (análogo al concepto de dificultad en los ítems de tests de rendimiento). Además, el modelo asume que todas las categorías de respuestas están ordenadas, en el sentido de que un sujeto no puede contestar una categoría de respuesta superior sin antes haber superado la categoría anterior.

Vistas las características fundamentales de este modelo, el objetivo de este trabajo es aplicar el MEC para obtener información sobre el funcionamiento de las categorías intermedias de este tipo de ítems, y, por tanto, de la relevancia de ellas en el ítem. Para ello se ha llevado a cabo un estudio donde se ha utilizado la Escala de Timidez de Cheek y Buss, (1981) en su versión revisada (Cheek, 1983) y traducida (Comeche, Díaz y Vallejo, 1995), donde se ha dividido al conjunto de sujetos en dos grupos en función de la información que proporcionan dichos sujetos sobre cómo han interpretado la categoría '*Ni de acuerdo ni en desacuerdo*': bien como una categoría que expresa una posición intermedia, bien como una categoría que expresa duda o indecisión. La información gráfica y numérica sobre las Curvas de Probabilidad de las Categorías de Respuestas de los ítems de la escala que nos proporciona el MEC nos permite analizar el comportamiento de dichas alternativas de respuestas en ambos grupos. A partir de estas curvas se definen, para el conjunto de los ítems, las denominadas 'regiones de respuestas más probables'. Estas regiones delimitan, en el continuo de la variable que se mide, distintas zonas en la que los sujetos responderán con mayor

probabilidad a una determinada categoría de respuesta. Lo que cabe esperar es que el MEC nos sirva para discriminar entre ambos grupos, de tal forma que en el grupo donde se ha interpretado '*Ni de acuerdo ni en desacuerdo*' como una categoría que expresa una posición intermedia, los pasos de las cinco categorías de respuesta de los ítems ( $\delta_{ij}$ ) deberían aparecer ordenadas en intensidad y, por tanto, definiría una 'región de respuesta más probable' para cada una de las cinco categorías. Por contra, en el grupo donde se ha interpretado '*Ni de acuerdo ni en desacuerdo*' como una categoría que expresa duda o indecisión, los pasos de las cinco categorías de respuesta de los ítems ( $\delta_{ij}$ ) no tendrían que aparecer ordenadas en intensidad, de forma que '*Ni de acuerdo ni en desacuerdo*' no definiría una 'región de respuesta más probable' para dicha categoría. En el primer caso, la categoría intermedia sería relevante y tendría sentido incluirla porque permite una medida más precisa en el continuo de la variable que se mide; en el segundo caso, la categoría intermedia sería irrelevante ya que no está proporcionando esa información, y quizás podría eliminarse.

## Método

### Sujetos

La muestra empleada en este trabajo estaba compuesta por un total de 185 sujetos, todos estudiantes universitarios. De ellos, un grupo de 146 (Grupo A) informaron que cuando respondieron a la categoría '*Ni de acuerdo ni en desacuerdo*' la utilizaron como si se tratara de una categoría intermedia o neutra. El grupo de los 39 restantes (Grupo B) informaron que respondieron a la misma categoría usándola como aquella a la que debían responder cuando tuviesen duda o no entendiesen el ítem.

### Materiales

El instrumento de medida utilizado en el estudio ha sido la 'Escala de Timidez' de Cheek y Buss (1981) en su versión revisada (Cheek, 1983) y traducida (Comeche, Díaz y Vallejo, 1995). Esta escala consta de 13 ítems con cinco alternativas de respuesta (1-Totalmente en desacuerdo; 2-En desacuerdo; 3-Ni de acuerdo ni en desacuerdo; 4-De acuerdo; y, 5-Totalmente de acuerdo).

En una página aparte se incluía una pregunta sobre cómo los sujetos habían entendido la alternativa de respuesta '3-Ni de acuerdo ni en desacuerdo', donde las opciones de respuestas eran dos: 1-Alternativa intermedia o neutra, y 2-Duda o incertidumbre sobre lo que dice el ítem.

La aplicación del Modelo de Escalas de Clasificación se ha efectuado mediante el programa desarrollado por Wright y Linacre (1992): BIGSTEPS ver. 2.29.

## Procedimiento

La aplicación de la escala se hizo de forma colectiva en las mismas aulas donde se impartían las clases. Una vez acabada la aplicación de la escala, todos los sujetos debían responder a un ítem en el que se preguntaba cómo habían entendido la categoría de respuesta '3-Ni de acuerdo ni en desacuerdo'.

## Resultados

### Ajuste

Sólo cuando el ajuste entre los datos y el modelo es adecuado se logran las ventajas que proporciona la utilización de la Teoría de Rasgo Latente en la construcción de tests (p.e. Wright y Stone, 1979; Hambleton y Swaminathan, 1985; Hambleton, 1990).

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, significan 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). Además, se incluyen el Residual Cuadrático Medio no ponderado Estandarizado (RCMnpE) y el Residual Cuadrático Medio ponderado Estandarizado (RCMpE), que no es más que el resultado de estandarizar los dos indicadores de ajuste anteriores. Estos estadísticos se distribuyen según una normal estandarizada, siempre y cuando no existan discrepancias entre los datos observados y los esperados por el modelo.

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 1 y 2), tanto para el grupo A como para el B. Los valores de RCMnpE y RCMpE también tienen medias cercanas a 0, aunque las desviaciones típicas no están tan cercanas a 1.

Tabla 1: *Ajuste global (ítems y sujetos) para el Grupo A.*

		RCMp	RCMpE	RCMnp	RCMnpE
SUJETOS	Medias	1,03	-0,1	1,02	-0,1
	D.Típicas	-----	1,5	-----	1,5
ÍTEMS	Medias	1,01	-0,1	1,02	0,1
	D.Típicas	-----	2,1	-----	2,0

Tabla 2: *Ajuste global (ítems y sujetos) para el Grupo B.*

		RCMp	RCMpE	RCMnp	RCMnpE
SUJETOS	Medias	1,07	0,0	1,05	0,0
	D.Típicas	-----	1,4	-----	1,4
ÍTEMS	Medias	1,00	0,0	1,02	0,2
	D.Típicas	-----	1,1	-----	1,3

A la vista de estos resultados, podemos afirmar que existe un ajuste más que aceptable del conjunto de datos (ítems y sujetos) al MEC.

### Calibración de los ítems y sus alternativas de respuesta

Los ítems de la 'Escala de Timidez' tienen todos cinco categorías de respuestas y por ello se han estimado cuatro parámetros para los pasos de cada ítem. Estos cuatro pasos o intersecciones entre categorías adyacentes vienen definidas por  $\delta_i$ , que es el valor de escala del ítem  $i$ , 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 de los ítems de los grupos A y B se recogen en la tabla 3 y 4, respectivamente. En estas tablas se muestran los valores de  $\delta_i$  (y de  $\tau_m$ ) de los todos los ítems (en *logits*).

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  $\tau_2$ , y así sucesivamente.

Tabla 3: *Calibración de los ítems y sus categorías de respuestas en el Grupo A.*

ÍTEMS	$\delta_i$	Primer Paso	Segundo Paso	Tercer Paso	Cuarto Paso
		$\delta_{i1}$	$\delta_{i2}$	$\delta_{i3}$	$\delta_{i4}$
1	-0,86	-3,30	-0,99	-0,76	1,61
2	0,09	-2,35	-0,04	0,19	2,56
3	0,00	-2,44	-0,13	0,10	2,47
4	0,70	-1,74	0,57	0,80	3,17
5	-0,01	-2,45	-0,14	0,09	2,46
6	-0,14	-2,58	-0,27	-0,04	2,33
7	0,09	-2,35	-0,04	0,19	2,56
8	-0,77	-3,21	-0,90	-0,67	1,70
9	-0,24	-2,68	-0,37	-0,14	2,23
10	0,76	-1,68	0,63	0,86	3,23
11	0,17	-2,27	0,04	0,27	2,64
12	0,03	-2,41	-0,10	0,13	2,50
13	0,18	-2,26	0,05	0,28	2,65

Parámetros de umbral:  $\tau_1 = -2,44$ ;  $\tau_2 = -0,13$ ;  $\tau_3 = +0,10$ ;  $\tau_4 = +2,47$

Tabla 4: Calibración de los ítems y sus categorías de respuestas en el Grupo B.

ITEMS	$\delta_i$	Primer Paso	Segundo Paso	Tercer Paso	Cuarto Paso
		$\delta_{i1}$	$\delta_{i2}$	$\delta_{i3}$	$\delta_{i4}$
1	-0,71	-3,28	-0,88	-0,89	2,21
2	0,23	-2,34	0,06	0,05	3,15
3	0,06	-2,51	-0,11	-0,12	2,98
4	0,45	-2,12	0,28	0,27	3,37
5	0,27	-2,30	0,10	0,09	3,19
6	0,10	-2,47	-0,07	-0,08	3,02
7	-0,41	-2,98	-0,58	-0,59	2,51
8	-0,44	-3,01	-0,61	-0,62	2,48
9	-0,28	-2,85	-0,45	-0,46	2,64
10	0,77	-1,80	0,60	0,59	3,69
11	0,10	-2,47	-0,07	-0,08	3,02
12	-0,11	-2,68	-0,28	-0,29	2,81
13	-0,03	-2,60	-0,20	-0,21	2,89

Parámetros de umbral:  $\tau_1=-2,57$ ;  $\tau_2=-0,17$ ;  $\tau_3=-0,18$ ;  $\tau_4=+2,92$

### Curvas Características de las Categorías de Respuestas

El MEC nos permite dibujar las curvas de probabilidad para cada categoría del ítem, siendo éstas idénticas para todos los ítems de la escala y difiriendo únicamente en la localización en el continuo de la variable que se mide ( $\delta_i$ ), que en nuestro caso es timidez. Las Curvas Características de las Categorías de Respuestas (CCCR) de un ítem son muy útiles para analizar el comportamiento de las alternativas de respuestas. Tanto es así, que podemos establecer para cada ítem las 'regiones de respuestas más probables' (Masters y Wright, 1997), que en el caso del MEC son idénticas en todos los ítems. Estas regiones definen en el continuo de adhesión distintas zonas en las que los sujetos responderán con mayor probabilidad a una determinada categoría de respuesta. Las interpretaciones de estas curvas y regiones, junto con los valores de adhesión de los pasos del ítem, tienen una gran importancia en el análisis de los ítems.

El ítem tipo de la Escala de Timidez, tanto para el Grupo A como para el B, representados en las figuras 1 y 2, tiene por categorías de respuestas las alternativas 'Totalmente en desacuerdo', 'En desacuerdo', 'Ni de acuerdo ni en desacuerdo', 'De acuerdo', y 'Totalmente de acuerdo' que están puntuadas con 1, 2, 3, 4 y 5 puntos, respectivamente; indicando que a mayor valor mayor aumento en la variable que mide el ítem (timidez). Pero mientras para el Grupo B, al fijarnos en la figura 2, vemos que siempre es menos probable responder a la categoría 'Ni de acuerdo ni en desacuerdo', (puntuada con un 3) frente a las categorías 'En desacuerdo' y 'De acuerdo', (puntuadas con un 2 y un 4 respectivamente).



das con 1 y 3, respectivamente); para el Grupo A (figura 1) esto no ocurre, y la categoría intermedia tiene un rango en el continuo en la que aparece como la más probable, lo que hace que para estos sujetos esta categoría sí sea relevante. O dicho de otro modo, los valores de los pasos de los ítems para el Grupo B no están ordenados, y en el Grupo A sí lo están (ver tablas 3 y 4).

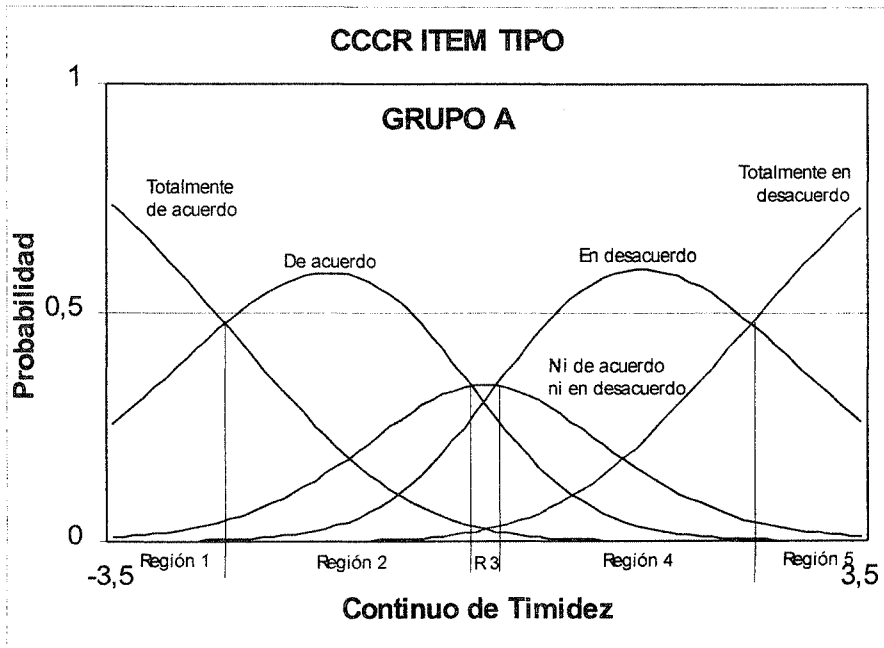


Figura 1: *Curvas Características de las Categorías de Respuestas del Ítem tipo del Grupo A.*

En estos casos, hemos podido comprobar que el MEC está, de alguna manera, poniendo de manifiesto que los sujetos han entendido la categoría de respuesta 'Ni de acuerdo ni en desacuerdo' con un significado realmente de categoría intermedia para el caso del grupo A, y con un significado más de duda o indecisión en el caso del Grupo B. Recordemos que todos los sujetos recibieron las mismas instrucciones iniciales, y sólo después de terminar la prueba informaron sobre su interpretación de la categoría 'Ni de acuerdo ni en desacuerdo'.

Además, a la luz de esta información, podemos plantearnos si la alternativa que hemos incluido cumple bien su función, de tal forma que si no produce una región de respuesta más probable, nos podríamos plantear bien eliminar la categoría

'Ni de acuerdo ni en desacuerdo' y dejar el ítem con las cuatro categorías restantes, bien modificar todas las alternativas, bien reformular todos los ítems, etc.

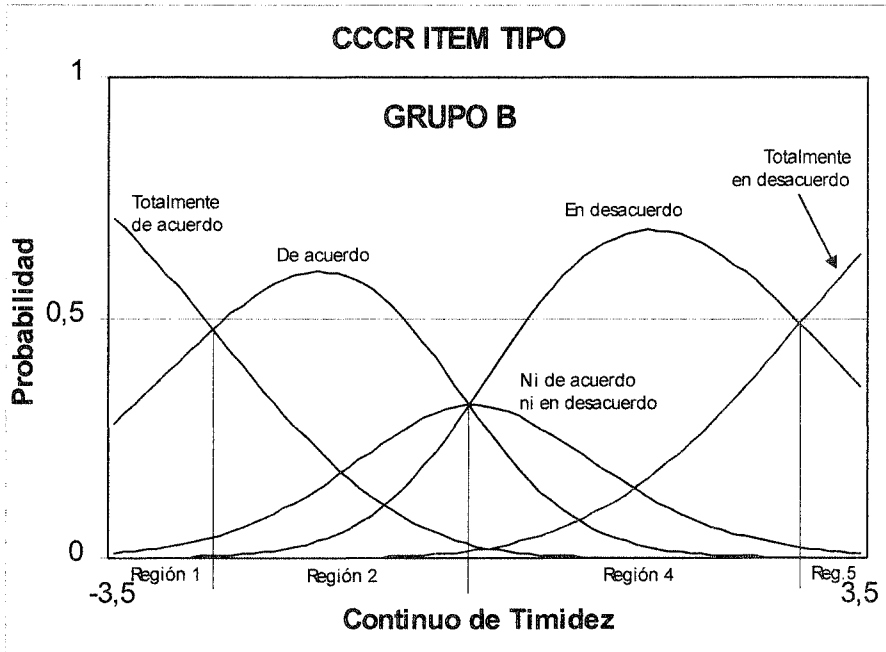


Figura 2: Curvas Características de las Categorías de Respuestas del Ítem tipo del Grupo B.

## Conclusiones

En el presente trabajo nos hemos centrado en el estudio de las alternativas intermedias y su relevancia y significado para los sujetos. No nos cabe duda que las CCCR y las regiones de respuestas más probables que nos proporciona el Modelo de Escalas de Clasificación pueden ayudarnos a tomar decisiones sobre estos aspectos referentes a las alternativas de respuesta 'Ni de acuerdo ni en desacuerdo'.

En el caso del Grupo A, que informaron responder a dicha categoría como si se tratara de una categoría intermedia, las CCCR producen una región de respuesta para esta alternativa. Ello no ocurre en el caso del Grupo B, que informaron responder a la misma categoría entendiéndola como duda o falta de entendimiento, donde las CCCR no originan dicha región de respuesta más probable. En el caso del Grupo A, tiene sentido incluir la categoría intermedia porque permite una medida más precisa

en el continuo de la variable que se mide; para el Grupo B, la categoría intermedia no está proporcionando esa información, y quizás debería eliminarse.

Como consideración general, podemos decir que si indicamos expresamente a los sujetos en las instrucciones que elijan lo menos posible esta categoría intermedia, solo cabe que la CCCR de esta categoría no defina una región de respuesta más probable; si, por contra, queremos que esa categoría sea una más de las demás dentro del ítem, esperaremos que dicha categoría intermedia defina una región de respuesta más probable. Pero quizás la clave esté en proporcionar unas indicaciones del test que especifiquen cómo funciona la categoría intermedia.

Sin duda la investigación que se presenta supone el comienzo de una línea de trabajo donde se pretende profundizar en el significado que tiene la categoría intermedia, por lo que será necesario ampliar estas investigaciones a otras escalas y con otras muestras.

## Referencias

- Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Andrich, D. (1978b). Scaling attitude items constructed and scored in the Likert Tradition. *Educational and Psychological Measurement*, 38, 665-680.
- Andrich, D. (1978c). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2 (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). Rating scales analysis. En J.P. Keeves (Ed.). *Educational research, methodology and measurement: an international handbook*. Elmsford, N.Y.: Pergamon Press, 297-303.
- Cañadas, I. y Sánchez, A. (1998). Categorías de respuestas en escalas tipo Likert. *Psicothema*, 10 (3), 623-631.
- Cheek, J.M. (1983). *The revised Cheek and Buss shyness scale*. Wellesley, Ma: Wellesley College. Texto no publicado.
- Cheek, J.M. y Buss, A.H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology*, 41, 330-339.

- Comeche, M.I., Díaz, M.I. y Vallejo, M.A. (1995). *Cuestionarios, inventarios y escalas. Ansiedad, depresión y habilidades sociales*. Madrid: Fundación Universidad-Empresa.
- Drasgow, F. (1995). Introduction to the polytomous IRT special issue. *Applied Psychological Measurement*, 19 (1), 1-3.
- DuBois, B. y Burns, J.A. (1975). An analysis of the question mark response category in attitudes scales. *Educational and Psychological Measurement*, 35, 869-884.
- Durán, A., Ocaña, A.C., Cañadas, I. y Pérez, F.J. (2000). Construcción de cuestionarios para encuestas: el problema de la familiaridad de las opciones de respuesta. *Metodología de Encuestas*, 2 (1), 27-60.
- Engelhard, G. y Stone, G. E. (1998). Evaluating the quality of ratings obtained from standard-setting judges. *Educational and Psychological Measurement*, 58 (2), 179-196.
- Glass, C.A.W. (1989). *Contributions to estimating and testing Rasch Models*. Doctoral dissertation, Technical University of Twente, Netherlands. The Hague: Cip-Gegevens Knoinglijke Bibliotheek.
- Hambleton, R.K. (1990). Item response theory: introduction and bibliography. *Psicothema*, 2 (1), 97-107.
- Hambleton, R.K. y Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer Academic Publishers.
- 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). Measurement models for ordered response categories. En R. Langeheine, y J. Rost, (Eds.). *Latent trait and latent class models*. Plenum Press, New York, 11-29.
- 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*, 101-121.
- 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*, 22 (1), 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*, 58, 87-99.
- Wright, B.D. y Linacre J.M. (1992). *A user's guide to BIGSTEPS*. Chicago. Mesa Press.
- Wright, B.D. y Masters, G.N. (1982). *Rating scale analysis*. Chicago. Mesa Press.
- Wright, B.D. y Stone, M.H. (1979). *Best test design: Rasch measurement*. Chicago. Mesa Press.

