

ESCALAS DE RESPUESTA TIPO LIKERT: ¿ES RELEVANTE LA ALTERNATIVA “INDIFERENTE”?

Ana Hernández Baeza
Begoña Espejo Tort
Vicente González Romá
Universitat de València

Juana Gómez Benito
Universitat de Barcelona

RESUMEN

En tests de actitudes y de personalidad es frecuente utilizar escalas de respuesta tipo Likert. En este tipo de escalas es frecuente también que se presente una categoría intermedia. Esta categoría representa una posición intermedia a lo largo del continuo y entre las categorías adyacentes. El presente estudio evalúa si las categorías de las escalas de respuesta de una serie de cuestionarios frecuentemente utilizados en la evaluación organizacional se presentan realmente ordenadas, tal y como asume el sistema de puntuación empleado. Asimismo, se evalúa si la categoría intermedia (“indiferente”) es relevante en términos de probabilidad de respuesta. Los cuestionarios analizados presentan una escala de respuesta tipo Likert de cinco puntos. Para lograr los objetivos propuestos, se ha ajustado el modelo nominal de Bock (1972). Los resultados obtenidos indican que si bien las categorías de respuesta aparecen ordenadas, la categoría intermedia “indiferente” no es relevante, es decir, en ningún intervalo de valores del rasgo latente aparece como la alternativa que presenta una mayor probabilidad de respuesta. Estos resultados sugieren eliminar la citada categoría de respuesta de las escalas de respuesta empleadas.

Palabras clave: *escalas Likert, categoría central, ordenación, relevancia, modelo nominal de Bock.*

Introducción

En las ciencias sociales, los tests y cuestionarios tienen una gran importancia por su aplicación en diversos ámbitos como son la educación, la orientación personal, la psicología clínica, etc. En los tests de actitudes y de personalidad es frecuente utilizar escalas de respuesta tipo Likert. Al responder a una escala Likert, los sujetos han de indicar su grado de acuerdo o desacuerdo con cada uno de los ítems de la escala. Para ello han de posicionarse en una escala de respuesta politómica con categorías ordenadas que oscilan entre el polo del acuerdo y el polo del desacuerdo. El número de categorías puede oscilar en las distintas escalas, siendo frecuente que se incluya una categoría intermedia como “indiferente”, “?” o “no estoy seguro”.

A cada una de las categorías de respuesta se le asignan sucesivos números enteros, de manera que reflejan ordinalmente el rasgo subyacente. Este sistema de puntuación se basa en la siguiente estructura relacional: el rasgo latente θ queda dividido en k intervalos mediante $k-1$ umbrales (τ), de manera que el número de umbrales sobrepasados por un sujeto desde el origen en el rasgo latente θ se corresponde con la puntuación observada en el ítem (y). En la figura 1 se presenta un ejemplo gráfico de este sistema de puntuación para un ítem con cinco categorías de respuesta.

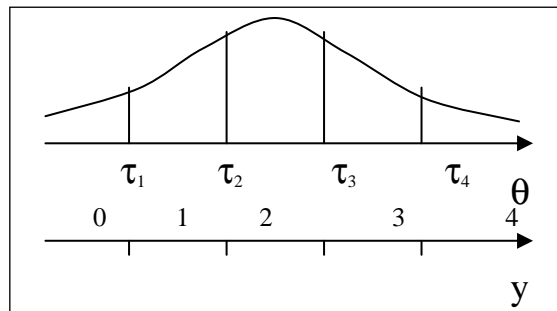


Figura 1: representación gráfica del sistema de puntuación en escalas tipo Likert.

Para el ejemplo de la figura 1,

$$y=0 \text{ si } \theta < \tau_1;$$

$$y=1 \text{ si } \tau_1 < \theta < \tau_2;$$

$$y=2 \text{ si } \tau_2 < \theta < \tau_3;$$

$$y=3 \text{ si } \tau_3 < \theta < \tau_4;$$

$$y=4 \text{ si } \theta > \tau_4,$$

A partir de este sistema de puntuación, se asume que la categoría intermedia representa una posición central respecto al resto de categorías a lo largo del continuo, implicando que los sujetos que seleccionan esta categoría lo hacen por ocupar en el

continuo una posición que está entre las que ocupan los sujetos que seleccionan las categorías adyacentes a la central.

Sin embargo, existe cierta controversia acerca del cumplimiento de este supuesto, poniéndose en tela de juicio el uso de las categorías centrales. Por una parte, ciertos estudios, llevados a cabo desde el marco de la Teoría Clásica de Tests, han recomendado el uso de las categorías intermedias (Edwards, 1957; Newman, 1979; Sudman y Bradburn, 1989), argumentando que es preferible no forzar las respuestas de los sujetos indecisos hacia un polo que podría no describirles. Sin embargo, otros estudios han cuestionado el uso de las categorías centrales porque éstas pueden atraer a las personas que las seleccionan por razones diferentes a la de su posición en el rasgo evaluado (Bock y Jones, 1968; Worthy, 1969; Kaplan, 1972; Dubois y Burns, 1975).

Algunas de las principales razones que podrían llevar a los sujetos a seleccionar las categorías centrales diferentes del significado asumido "sentirse en el punto medio" son la ambivalencia y la indiferencia (Cronbach, 1946; Edwards, 1946; Worthy, 1969; Goldberg, 1971; Kaplan, 1972; DuBois y Burns 1975). El sujeto ambivalente es el que presenta sentimientos positivos y negativos acerca de la misma cuestión, por lo que la ambivalencia sería resultado de una alta implicación con el objeto de evaluación. El sujeto indiferente, en cambio, es aquél que no está interesado por la cuestión, siendo resultado de una baja implicación con el objeto de evaluación. Otras razones señaladas en la literatura son la existencia de estilos de respuesta caracterizados por escoger la categoría intermedia (Worthy, 1969; Dubois y Burns, 1975), la falta de comprensión del enunciado (Dubois y Burns, 1975), la falta de competencia o información relativa al enunciado (Dubois y Burns, 1975), el recelo a revelar cuestiones personales (Dubois y Burns, 1975) y la falta de familiaridad o baja frecuencia de uso de los ítems empleados en una lista de adjetivos (Cruickshank, 1984).

En este sentido, Espejo, González-Romá e Hidalgo (1999) analizaron el significado atribuido a tres categorías intermedias diferentes empleadas en muchos de los ítems del 16PF (Cattell, 1972) y del 16 PF-5 (Cattell, Cattell y Cattell, 1995). Concretamente, en el 16 PF se utilizan las categorías intermedias "término medio" y "no estoy seguro", mientras que en el 16 PF-5 se utiliza en todos los ítems la categoría intermedia "?". Los resultados del estudio mostraron que aproximadamente en un 50% de las ocasiones en que se escogió alguna de las categorías intermedias el significado que se atribuyó a estas categorías fue diferente al de estar en el punto medio. Algunas de las razones más frecuentemente aportadas fueron indicadores de ambivalencia y de falta de información para decantarse por uno de los polos (en torno al 12% de las ocasiones en cada caso).

Los significados diferentes a estar en el punto medio atribuidos a la categoría central amenazan el supuesto de unidimensionalidad (Cheung y Mooi, 1994; Andrich, Jong y Sheridan, 1997). Así, por ejemplo, si un sujeto escoge la categoría intermedia de un ítem que mide extraversión porque no entiende su significado, su respuesta será independiente de su nivel de extraversión. Esta violación del supuesto de unidimensionalidad podría dar lugar a una ordenación incorrecta de las categorías de respuesta (Cheung y Mooi, 1994; Andrich y otros, 1997), lo que implicaría la

ruptura de la relación monotónica entre las categorías y el rasgo latente subyacente. Así pues, el cumplimiento del supuesto del orden de las categorías resulta esencial tanto para obtener medidas precisas del rasgo como para que las inferencias realizadas a partir de dichas medidas resulten válidas.

Sin embargo, a pesar de que el supuesto del orden de las categorías puede ser puesto a prueba desde el marco de la Teoría de Respuesta a los Items, dicho supuesto apenas ha sido evaluado empíricamente. Una excepción es el estudio de Espejo y González-Romá, (1999), que evaluaron el orden de las alternativas de respuesta en las escalas de tres alternativas de diferentes factores del Cuestionario de personalidad 16PF (Cattell, 1972) con categorías intermedias “no estoy seguro” y “término medio”. Dicho funcionamiento fue evaluado a partir de las estimaciones ofrecidas por el modelo nominal de Bock (Bock, 1972). Los resultados mostraron que de los factores analizados (los factores C y F, con la categoría intermedia “no estoy seguro”, y los factores E, G, H, L, O y Q4 con la categoría intermedia “término medio”), sólo los factores H y Q4 mostraron las categorías de respuesta ordenadas. Sin embargo, incluso cuando las categorías sí estuvieron ordenadas, la relevancia de la categoría intermedia quedaba puesta en tela de juicio, ya que, a partir de las estimaciones de los umbrales entre categorías adyacentes, se pudo observar que la probabilidad de seleccionar la categoría intermedia nunca era superior a la probabilidad de elección de cualquiera de las otras categorías, independientemente de la posición de los sujetos en el continuo latente.

Por otra parte, Andrich y otros (1997) evaluaron un cuestionario de actitudes de los profesores hacia el uso de estrategias de instrucción directa. Dicha actitud fue medida antes y después de que los profesores fueran expuestos a la estrategia en cuestión. El cuestionario debía responderse mediante una escala Likert de cinco alternativas, con categoría central “no estoy seguro”. En este caso las alternativas de respuesta fueron evaluadas mediante el Modelo de Escalas de Clasificación (Rating Scale Model) propuesto por Andrich (1978) (ver también Rojas, 1999; Rojas y Hernández, 2000). Puesto que dicho modelo asume el orden entre las categorías de respuesta, la evaluación de las mismas se centró en los umbrales entre cada par de categorías adyacentes. Los resultados mostraron que dichos umbrales no siempre aparecían ordenados. En concreto, 8 de 40 ítems en el pre-test, y 19 de 40 ítems en el post-test presentaron umbrales no ordenados. En estos casos, de nuevo, la categoría intermedia “no estoy seguro” resultó ser la principal responsable de la desordenación de los umbrales entre las categorías de respuesta adyacentes, siendo la probabilidad de seleccionar la categoría central inferior a la probabilidad de seleccionar cualquiera de las otras categorías para todos los niveles del rasgo. En una segunda parte del estudio se eliminó la posibilidad de seleccionar la categoría intermedia y el número de ítems con umbrales desordenados se redujo a uno. Este último resultado es congruente con los resultados obtenidos por Andrich y Schoubroeck (1989) con la versión de 30 ítems del cuestionario GHQ (General Health Questionnaire) (Goldberg, 1978). Los ítems de este cuestionario presentan escalas de respuesta con cuatro categorías, sin categoría central. Los resultados obtenidos mostraron que todos los ítems presentaban umbrales ordenados.

En el ámbito español, Rojas y Fernández (2000) evaluaron el funcionamiento de la categoría central “ni de acuerdo ni en desacuerdo” en dos grupos de sujetos definidos en función de cómo habían interpretado dicha categoría: como una categoría que expresa una posición intermedia en el continuo, o como una categoría que expresa duda o indecisión. En ambos grupos una escala de respuesta tipo Likert de cinco alternativas fue evaluada mediante el Modelo de Escalas de Clasificación. Los resultados mostraron que, mientras para el primer grupo la categoría central presentaba una región del continuo para la que era más probable escoger dicha categoría frente a las otras, en el segundo grupo la categoría central no resultó relevante. Es decir, nunca presentó una probabilidad de elección superior a la de las otras categorías. Los resultados de este estudio apoyarían la idea apuntada por diferentes autores acerca de que los diferentes significados otorgados a la categoría intermedia, distintos del asumido “sentirse en el medio del continuo”, son la principal causa de que la categoría intermedia no funcione como tal.

Todos estos resultados sugieren que la categoría intermedia es la que resulta especialmente problemática, ya que o bien rompe el orden entre las categorías o bien no resulta ser relevante. Teniendo en cuenta la escasez de estudios en los que, además de la relevancia, se haya evaluado el orden de las alternativas de respuesta, el objetivo de este trabajo es determinar tanto si se mantiene el orden entre las categorías en escalas de respuesta que incluyen la categoría “indiferente”, como determinar si dicha categoría resulta relevante (esto es, si en alguna zona del continuo del rasgo latente, la categoría “indiferente” es la alternativa de respuesta con mayor probabilidad de elección). En concreto se emplean escalas tipo Likert de cinco alternativas, frecuentemente utilizadas en diferentes ámbitos de las ciencias sociales.

Método

Muestra

La muestra del presente estudio está compuesta por 932 empleados de un servicio público de salud que participaron en una investigación sobre calidad de vida laboral. El procedimiento de muestreo llevado a cabo para su selección fue bietápico. En la primera etapa se seleccionaron aleatoriamente 250 equipos de trabajo. En la segunda etapa se seleccionaron cuatro sujetos por equipo; tres de ellos fueron seleccionados aleatoriamente, siendo el cuarto el líder del equipo. Del total de la muestra, el 37,9% fueron varones y el 58,3% mujeres. La media de edad de la muestra fue de 41,18 años (d.t. 9,18). Por lo que se refiere a la antigüedad en el sistema público de salud, el promedio fue de 14,58 años (d.t. 7,9). Finalmente, por lo que se refiere al grupo ocupacional, un 4,1% de la muestra eran directivos, un 22,4% médicos, un 32,9% enfermeras, un 14,8% era personal sanitario no facultativo (p.e., fisioterapeutas), un 8,3% era personal de administración, un 14,3% personal de mantenimiento, y un 2,3% eran técnicos no facultativos (p.e., ingenieros).

VARIABLES E INSTRUMENTOS DE MEDIDA

Para lograr el objetivo del presente trabajo, se ha analizado un conjunto de ítems de diferentes cuestionarios diseñados para medir tres constructos frecuentemente recurridos en estudios de calidad de vida laboral en las organizaciones: el clima de apoyo, la sobrecarga de rol y la satisfacción laboral con el puesto de trabajo. Por lo que respecta al clima de apoyo, éste hace referencia a la calidad percibida de las relaciones interpersonales en el trabajo, es decir, al grado en que los compañeros muestran interés por los problemas de los otros y se ayudan mutuamente. La sobrecarga de rol se refiere al grado en que el individuo percibe que tiene una cantidad de trabajo excesiva o que le falta tiempo para realizar dicho trabajo adecuadamente. Por último, la satisfacción laboral con el puesto de trabajo indica en qué grado el individuo está satisfecho con los recursos que cuenta para realizar su trabajo, las actividades que desarrolla, y las oportunidades de formación y de realización personal que le ofrece su puesto de trabajo.

Las distintas variables fueron medidas mediante los ítems incluidos en una batería diseñada, en el marco de un estudio más amplio, para medir diferentes aspectos de la calidad de vida laboral (Peiró y González-Romá, 1994). Específicamente, la variable “clima de apoyo” fue medida mediante tres ítems; la variable “satisfacción laboral con el puesto” fue medida mediante cuatro ítems y, finalmente, la variable “sobrecarga de rol” fue medida mediante los tres ítems que componen el cuestionario de sobrecarga de rol de Camman, Fichman, Jenkins y Klesh (1979). En la tabla 1 se muestran los estadísticos y la consistencia interna de cada subescala, y en el anexo 1 se presentan los ítems empleados en los análisis. En los cuestionarios de clima de apoyo y sobrecarga de rol se utilizó una escala de respuesta tipo Likert que oscilaba entre “muy en desacuerdo” y “muy de acuerdo”, mientras que en el cuestionario de satisfacción con el puesto de trabajo la escala oscilaba entre “muy insatisfecho” y “muy satisfecho”. En todos los casos se presentaron cinco alternativas de respuesta, siendo la categoría intermedia “indiferente”.

Tabla 1: *estadísticos descriptivos y consistencia interna (α) de cada subescala.*

	Núm. ítems	Media	d.t.	α
Clima de apoyo	3	3,98	0,96	0,86
Sobrecarga de rol	3	3,27	0,82	0,67
Satisfacción laboral con el puesto	4	3,05	1,1	0,83

ANÁLISIS

La unidimensionalidad de cada uno de los instrumentos de medida se analizó mediante análisis factorial de componentes principales. Se concluyó que existía

evidencia empírica que apoyaba la unidimensionalidad cuando el cociente entre el valor propio del primer componente y el valor propio del segundo componente era elevado (Reckase, 1979; Lord, 1980), y la representación gráfica de los valores propios asociados a cada componente indicaba la existencia de un factor dominante (Hulin, Drasgow y Parsons, 1983; Hambleton y Swaminathan, 1985).

Tras evaluar la unidimensionalidad, el modelo nominal de Bock (1972, 1997) fue ajustado a los datos mediante el programa Multilog 6.0 (Thissen, 1991). Bajo este modelo, la probabilidad de seleccionar cada una de las categorías de respuesta (k) de un ítem (j), dado un determinado nivel de rasgo (θ), se expresa en los siguientes términos:

$$P_{jk}(\theta) = \exp(\theta a_{jk} + c_{jk}) / \sum_{k=1}^K \exp(\theta a_{jk} + c_{jk})$$

Así pues, la probabilidad de seleccionar una determinada categoría de un ítem, dado un determinado nivel de rasgo, es función de dos parámetros para cada una de las categorías de respuesta: a_{jk} y c_{jk} . El parámetro a_{jk} es el parámetro de discriminación de cada categoría de respuesta para cada ítem y c_{jk} es el logaritmo natural de la razón entre la probabilidad de seleccionar la categoría k y la probabilidad de seleccionar la primera categoría cuando θ es igual a cero (Mellenbergh, 1995). Si $c_{jk} > 0$, para $\theta=0$, la probabilidad de preferir la categoría k será mayor que la probabilidad de preferir la primera categoría; si $c_{jk} < 0$, para $\theta=0$, la probabilidad de preferir la categoría k será menor que la probabilidad de preferir la primera categoría.

Este modelo, que no asume ningún orden entre las categorías de respuesta, permite comprobar si las estimaciones del parámetro a_{jk} se presentan ordenadas. Samejima (1972) mostró que cuando los parámetros a_{jk} están ordenados, las probabilidades acumuladas de las sucesivas categorías lo están también. Si la probabilidad de seleccionar una categoría superior aumenta al aumentar los niveles de θ , las funciones de respuesta de las categorías estarán ordenadas, lo que se reflejará en la ordenación de los parámetros a de cada categoría. En este caso, se podrá concluir que las categorías de respuesta están ordenadas para ese ítem.

Una vez comprobado el orden de las categorías, también se puede determinar si todas las categorías de respuesta son relevantes. Esta relevancia puede determinarse a partir de las estimaciones de los umbrales entre las categorías de respuesta consecutivas. Cuando dichos umbrales se presenten ordenados se podrá concluir que todas las categorías de respuesta son relevantes, es decir, que todas ellas tienen un rango de θ para el cual tienen la mayor probabilidad de ser seleccionadas frente al resto de categorías. Los umbrales entre cada par de categorías consecutivas (τ_{jk}) pueden calcularse a partir de la siguiente expresión (Bock, 1972; 1997):

$$\tau_{jk} = (c_{jk} - c_{j,k-1}) / (a_{j,k-1} - a_{jk})$$

Ambos aspectos, orden y relevancia de las categorías, se tienen en cuenta en el presente estudio para evaluar el funcionamiento de las escalas de respuesta.

Resultados

Los análisis de la unidimensionalidad de los instrumentos de medida indicaron que en todos ellos existía un solo factor dominante. En el caso de “clima de apoyo” el primer componente mostró un valor propio de 2,34 (78.1% de varianza explicada), y el segundo de 0,37 (12,4% de varianza explicada). Para la “satisfacción con el puesto de trabajo”, los dos primeros componentes presentaron valores propios de 2,08 (52% de varianza explicada) y 0,78 (19,5% de varianza explicada), respectivamente. Por último, para la variable ‘sobrecarga de rol’ el primer componente mostró un valor propio de 2,23 (74,6% de varianza explicada), y el segundo de 0,42 (14% de varianza explicada).

En las tablas 2a y 2b se muestran los resultados obtenidos tras ajustar el modelo nominal de Bock (1972). Como se puede observar en la segunda columna de la tabla, todos los modelos presentaron un buen ajuste, ya que el índice G^2/gl fue menor a 3 en todos los casos (Drasgow, Levine, Tsien, Williams y Mead, 1995). Por consiguiente, las estimaciones ofrecidas por el modelo fueron empleadas para evaluar el orden y la relevancia de las categorías de respuesta.

Tabla 2a: Índices de bondad de ajuste (G^2/gl) y parámetros estimados para cada categoría de respuesta (k) de los ítems 1 y 2 de las escalas analizadas.

SATISFACCIÓN CON EL PUESTO	G^2/gl	Ítem 1			Ítem 2		
		a	C	τ	a	c	τ
k1	0,95	-1	0,10	-1,49	-7,04	-12,87	-2,27
k2		-0,45	0,92	5,77	-0,77	1,35	-1,29
k3		-0,32	0,17	-0,89	0,60	3,12	-0,86
k4		0,30	0,72	2,25	2,54	5,23	0,97
k5		1,47	-1,91		4,66	3,17	
SOBRECARGA	G^2/gl	a	C	τ	a	c	τ
k1	2,91	-2,96	-1,49	-2,46	-5,27	-3,91	-1,35
k2		-1,98	0,92	0,22	-1,72	0,89	-0,32
k3		-0,25	0,54	-0,52	0,32	1,54	-0,12
k4		1,30	1,35	-1,42	2,45	1,79	1,19
k5		3,18	-1,32		4,21	-0,31	
CLIMA APOYO	G^2/gl	a	c	τ	a	c	τ
k1	2,63	-3,43	-4,76	-2,20	-6,81	-9,69	-1,85
k2		-1,69	-0,93	1,03	-1,92	-0,62	-1,38
k3		-0,43	0,37	-1,38	-0,62	1,17	-0,95
k4		1,37	2,85	0,14	3,14	4,76	0,12
k5		4,18	2,47		6,20	4,38	

Por lo que se refiere al orden, los parámetros a aparecieron ordenados para todos los ítems, lo que indica que las categorías de respuesta correspondientes (k) están ordenadas en todos los casos. En cuanto a la relevancia, los resultados mostraron que el orden entre los umbrales de las categorías de respuesta consecutivas (τ) no se mantenía en todos los casos. Esto ocurre concretamente para tres de los cuatro ítems de la escala de satisfacción con el puesto de trabajo (ítems 1, 3 y 4), para dos de los tres ítems de sobrecarga (ítems 1 y 3), y para uno de los tres ítems de apoyo (ítem 1) (ver tabla 2a y b). En todos estos casos, la categoría intermedia estuvo implicada, ya que se desordenaba o bien el umbral entre las categorías 2 y 3, o bien el umbral entre las categorías 3 y 4. Esto indica que la categoría intermedia es la categoría que aparece como no relevante, de modo que no presenta una mayor probabilidad de ser escogida frente a cualquier otra alternativa de respuesta para ningún valor del rasgo.

Tabla 2b: Índices de bondad de ajuste (G^2/gl) y parámetros estimados para cada categoría de respuesta (k) de los ítems 3 y 4 de las escalas analizadas.

		Ítem 3			Ítem 4		
SATISFACCIÓN CON EL PUESTO	G^2/gl	A	C	τ	a	c	τ
k1		-2,42	-2,94	-1,58	-1,21	0,42	-0,30
k2		-0,89	-0,52	-1,28	-0,59	0,63	1,09
k3	0,95	-0,06	0,54	-1,68	-0,36	0,38	-0,32
k4		0,84	2,05	0,69	0,20	0,56	1,45
k5		2,54	0,87		1,96	-1,99	
		Ítem 3			Ítem 4		
SOBRECARGA	G^2/gl	a	C	τ			
k1		-3,20	-0,60	-0,86			
k2		-1,15	1,16	0,76			
k3	2,91	-0,11	0,37	-0,35			
k4		1,28	0,86	1,39			
k5		3,19	-1,79				
		Ítem 3			Ítem 4		
CLIMA APOYO	G^2/gl	a	C	τ			
k1		-3,09	-4,25	-2,28			
k2		-1,68	-1,03	-1,41			
k3	2,63	-0,41	0,76	-1,14			
k4		1,26	2,67	0,31			
k5		3,91	1,85				

En la figura 2 se presentan, a modo de ejemplo, las curvas de las opciones de respuesta de un ítem en el cual la categoría intermedia es irrelevante (ítem 3 de la escala de sobrecarga de rol). Como puede observarse, las opciones de respuesta sí aparecen ordenadas. Sin embargo, no todas las categorías de respuesta son relevantes. En concreto, la categoría intermedia (k3) no presenta una mayor probabilidad de

ser seleccionada que cualquiera de las otras categorías en ningún intervalo de valores del rasgo latente. Los sujetos con niveles medios en el rasgo tienen más probabilidad de seleccionar las categorías 2 y 4 que la categoría central “indiferente”.

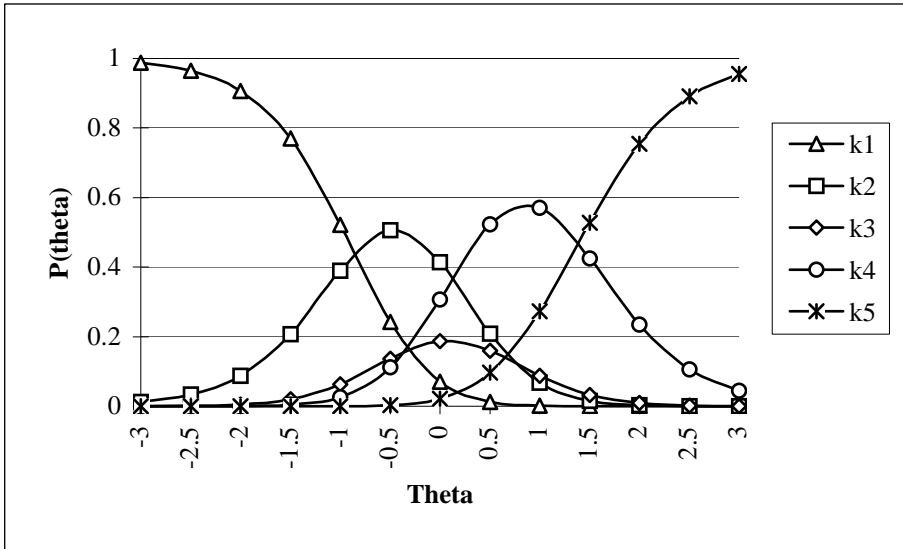


Figura 2: curvas de las categorías de respuesta (k) para el ítem 3 de la escala de sobrecarga de rol.

El resto de los ítems (ítem 2 de satisfacción con el puesto, ítem 2 de sobrecarga de rol e ítems 2 y 3 de clima de apoyo) sí presentaron los umbrales de respuesta ordenados, por lo que todas las categorías de respuesta podrían considerarse relevantes. Sin embargo, cabe destacarse que en todos estos casos, el rango de valores del rasgo en el que la categoría intermedia presentó la mayor probabilidad de ser escogida fue muy pequeño. A continuación se presenta, a modo de ejemplo, la gráfica de las curvas para las alternativas de respuesta de uno de estos ítems (ítem 3 de la escala de clima de apoyo) (ver figura 3).

Como se ve en la gráfica, hay un rango de valores en el rasgo para los cuales la probabilidad de escoger la categoría intermedia es mayor que la probabilidad de seleccionar cualquier otra alternativa de respuesta. Sin embargo, es un rango muy restringido de valores. En concreto, para el ítem representado, este rango de valores oscila entre $-0,95$ y $-1,38$.

Discusión

El uso de las categorías centrales ha sido cuestionado en la literatura por diferentes autores (DuBois y Burns, 1975; Andrich y otros, 1997; Espejo y otros, 1999; Espejo y González–Romá, 1999), ya que, en ocasiones, la selección de la categoría intermedia podría ser independiente del nivel que presentan los sujetos en el rasgo latente. Esto violaría el supuesto de unidimensionalidad, lo que podría dar lugar a una desordenación de las categorías de respuesta y/o de los umbrales entre las mismas. A partir de aquí el objetivo del presente trabajo se ha centrado en determinar si se mantiene el orden entre las categorías en escalas de respuesta que incluyen la categoría central “indiferente”, así como determinar si ésta resulta relevante, es decir, si existe un intervalo de valores del continuo del rasgo latente en el que la categoría “indiferente” sea la alternativa de respuesta con mayor probabilidad de elección. Para ello, se evaluó un conjunto de ítems que medían diferentes constructos ampliamente utilizados y estudiados en contextos organizacionales (satisfacción laboral con el puesto de trabajo, sobrecarga de rol y clima de apoyo). En todos los casos se empleó una escala de respuesta tipo Likert de cinco alternativas, en la que aparecía la categoría de respuesta “indiferente” como categoría central.

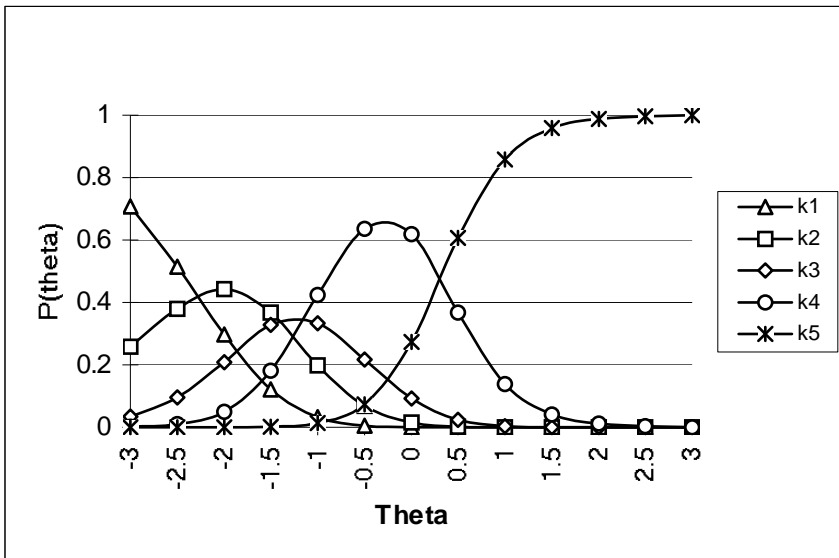


Figura 3: curvas de las categorías de respuesta (k) para el ítem 3 de la escala de clima de apoyo.

Por lo que se refiere al orden, los resultados mostraron que, en todos los casos, las categorías de respuesta aparecían ordenadas. Sin embargo, no ocurrió lo mismo con los umbrales entre las categorías de respuesta adyacentes. Para las tres variables evaluadas este desorden se manifestó en seis de los ítems evaluados, para los cuales la categoría intermedia no resultó ser relevante. Para los cuatro ítems restantes, si bien los umbrales sí aparecieron ordenados, el intervalo de valores en el rasgo para los cuales la categoría central presentó la mayor probabilidad de ser seleccionada frente al resto de categorías, fue muy pequeño.

A partir de estos resultados se puede concluir que en las escalas estudiadas, con cinco alternativas de respuesta y categoría central “indiferente”, se mantiene el orden de las alternativas de respuesta. Estos resultados contrastan con los obtenidos por Espejo y otros (1999). Estos autores estudiaron las categorías centrales “no estoy seguro” y “término medio” en escalas supuestamente ordinales de tres alternativas y observaron que era frecuente que las categorías de respuesta no se presentaran ordenadas. Así pues, parece que el número de categorías de respuesta y los correspondientes anclajes verbales, pueden afectar al cumplimiento del supuesto de orden entre las mismas.

Por lo que se refiere a la relevancia de la categoría central “indiferente”, los resultados mostraron que dicha categoría o bien nunca tiene mayor probabilidad de ser seleccionada que el resto de categorías, independientemente de cuál sea el nivel de rasgo de los sujetos, o bien esta mayor probabilidad de elección se da en un rango muy restringido de valores en el rasgo subyacente. Estos resultados son congruentes con los obtenidos por Andrich y otros (1997) y Espejo y otros (1999). En concreto, Andrich y otros (1997) observaron que para ciertos ítems la categoría central “no estoy seguro” no era relevante. Asimismo, Espejo y otros (1999) observaron que, para los ítems que presentaron las categorías de respuesta ordenadas, la alternativa central “término medio” tampoco era relevante.

Todos estos resultados muestran que, incluso para las personas que ocupan una posición intermedia en el continuo subyacente, y que por lo tanto cabría esperar que seleccionasen con mayor probabilidad la categoría intermedia, resulta más probable que elijan las categorías adyacentes a ésta. Consecuentemente, la categoría central (categoría 3) no es una respuesta típica de los sujetos con niveles medios en el rasgo, lo que puede afectar tanto a la precisión de las medidas como a la validez de las inferencias realizadas. Los resultados del presente estudio, junto con los obtenidos previamente por Espejo y González-Romá (1999), y por Andrich y otros (1997) sugieren que la falta de relevancia de la categoría central no es infrecuente, ya que estos efectos se han observado independientemente del anclaje verbal utilizado para la alternativa intermedia, e independientemente del número de anclajes empleados (tres o cinco). Todos estos resultados cuestionan la utilización de las categorías centrales en escalas de respuesta tipo Likert.

El hecho de que la categoría central no presente una alta probabilidad de elección para la gente que ocupa una posición intermedia en el continuo, es decir, su falta de relevancia, podría deberse a varios factores. En primer lugar, a que el significado otorgado a la categoría central sea diferente del asumido “sentirse en el medio”, ya que en este caso se esperaría una relativa independencia entre el nivel de

rasgo mostrado por los sujetos y la probabilidad de elección de dicha categoría. Esta explicación es congruente con los resultados obtenidos por Rojas y Fernández (2000), puesto que, en este estudio, cuando los sujetos interpretaron la alternativa intermedia como una categoría que expresa duda o indecisión, dicha categoría no resultó ser relevante. Otro factor a considerar es la falta de familiaridad o baja frecuencia de uso de los anclajes verbales empleados para definir las posiciones centrales del continuo (Durán, Ocaña, Cañadas y Pérez, 2000). Por último, cabe considerar la posible tendencia de los sujetos encuestados a responder con actitudes ligeramente extremas, y por tanto a evitar la categoría central, por efecto de la aquiescencia. La influencia de estos posibles factores en la falta de relevancia de la categoría central debería ser evaluada en posteriores estudios.

Por otra parte, puesto que se ha visto que ciertos ítems presentan ordenados tanto las categorías de respuesta como los umbrales entre las mismas (es decir, todas las categorías están ordenadas y son relevantes), sería conveniente realizar estudios comparativos de escalas de respuesta con y sin categoría intermedia. Esto permitiría determinar si la eliminación de la alternativa central implica una pérdida relevante de precisión y de validez, en cuyo caso habría que buscar soluciones alternativas a su eliminación. Si la pérdida de validez y precisión no es sustancial, la eliminación de la categoría intermedia sería la alternativa más razonable.

Finalmente cabe señalar ciertas limitaciones del presente estudio. En primer lugar, el número de ítems que formaban los cuestionarios analizados era reducido (tres y cuatro). En segundo lugar sólo se ha evaluado la categoría central "indiferente" en escalas tipo Likert. Otros anclajes verbales, así como otros formatos de respuesta más objetivos (como escalas de frecuencia) deberían ser evaluados en futuros estudios. Asimismo, puesto que los cuestionarios empleados se han restringido a contextos organizacionales, se deberían tener en cuenta otros ámbitos de aplicación de este tipo de cuestionarios.

Referencias

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Andrich, D.; Jong, J. y Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. En J. Rost y R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences*. Münster, Germany: Waxman Verlag. pp. 58-68.
- Andrich, D. y Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, 19, 469-485.
- Bock, R. y Jones, L. V. (1968). *The measurement and prediction of judgement and choice*. San Francisco: Holden Day.
- Bock, R. D. (1972). Estimating item parameters and latent ability when the responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.

- Bock, R. D. (1997). The nominal categories model. En W. J. van der Linden y R. K. Hambleton (eds.): *Handbook of modern item response theory*. New York Inc.: Springer-Verlag.
- Camman, C.; Fichman, M., Jenkins, D. y Klesh, J. (1979). *The Michigan Organizational Assessment Questionnaire*. Manuscrito no publicado, University of Michigan, Ann Arbor.
- Cattell, R. B. (1972). [16 PF Cuestionario de personalidad (adolescentes y adultos)]. Traducido y adaptado por TEA ediciones. Publicaciones de Psicología Aplicada. Serie menor nº 89.
- Cattell, R. B.; Cattell, A. K. S y Cattell, H. E. P (1995). [16 PF-5]. Traducido y adaptado por TEA ediciones. Publicaciones de Psicología Aplicada. Serie menor nº 228.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475-494.
- Cruickshank, P. J. (1984). A stress and arousal mood scale for low vocabulary subjects: a reworking of Mackay et al. (1978). *British Journal of Psychology*, 75, 89-94.
- Cheung, K. C. y Mooi, L. C. (1994). A comparison between the rating scale model and dual scalind for Likert scales. *Applied Psychological Measurement*, 18, 1-13.
- Drasgow, F., Levine, M. V., Tsien, S., Williams, B, y Mead, A. D. (1995). Fitting polytomous item response theory models to multiple choice tests. *Applied Psychological Measurement*, 19, 143-165.
- Dubois, B. y Burns, J. A. (1975). An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement*, 35, 869-884.
- Durán, A.; Ocaña, A. C.; Cañadas, I. y Pérez Santamaría, 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.
- Edwards, A. L. (1946). A critique of "neutral" items in attitude scales constructed by the method of equal appearing intervals. *Psychological Review*, 53, 159-169.
- Edwards, A. L. (1957). *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts.
- Espejo, B.; González-Romá, V. y Hidalgo, M. D. (1999). El orden de las alternativas de respuesta en escalas tipo Likert: un estudio mediante modelos de la Teoría de Respuesta al Ítem. Trabajo presentado al VI Congreso de Metodología de las CC. Sociales y de la Salud. Oviedo, España
- Espejo, B. y González-Romá, V. (1999). *El significado de las categorías centrales en las escalas tipo Likert*. Trabajo presentado al VI Congreso de Metodología de las CC. Sociales y de la Salud. Oviedo, España.
- Goldberg, D. P. (1978). *Manual of the General Health Questionnaire*. National Foundation for Educational Research: Windsor.
- Goldberg, G. (1971). *Response format in attitude scales*. Manuscrito no publicado, Northwestern University.

- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory: A suggested modification of the semantic differential technique. *Psychological Bulletin*, 77, 361-372.
- Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, 19, 91-100.
- Neumann, L. (1979). Effects of categorization on relationships in bivariate on relationships in bivariate distributions and applications to rating scales. Dissertation Abstracts International, 40, 2262-B
- Peiró, J. M. y González-Romá, V. (1994). *Estudio sobre los correlatos psicológicos del absentismo en el personal del Servicio Valenciano de Salud*. Valencia: Consellería de Sanitat i Consum de la Generalitat Valenciana.
- Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics*, 4 (3), 207-230
- Rojas, A. J. (1999). Aplicación del modelo de escalas de clasificación para la medición de la "acción política no convencional". *Metodología de Encuestas*, 1 (1), 1-18.
- Rojas, A. J. y Fernández, J. S. (2000). Análisis de las alternativas de respuestas intermedias mediante el modelo de escalas de clasificación. *Metodología de Encuestas*, 2 (2), 171-183.
- Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No. 18.
- Sudman, S. y Bradburn, N. M. (1989). Measuring attitudes: recording responses. En Sudman, S. y Bradburn, N. M. (Eds.). *Asking questions: a practical guide to questionnaire design*. San Francisco: Jossey-Bass Publishers.
- Thissen, D. (1991). MULTILOG user's guide. Multiple, categorical item analysis and test scoring using item response theory (Version 6.0). Computer program. Chicago, IL: Scientific Software.
- Worthy, M. (1969). Note on scoring midpoint responses in extreme response style scores. *Psychological Reports*, 24, 189-190.

Anexo 1

Satisfacción laboral con el puesto de trabajo

Indique el grado de satisfacción o insatisfacción que le producen cada uno de los siguientes aspectos relacionados con su trabajo:

1. Los medios y recursos de que dispone para realizar su trabajo.
2. Las tareas y actividades que usted realiza en su jornada de trabajo.
3. La realización personal que usted consigue en su trabajo.
4. Las oportunidades de formación que le ofrece su trabajo.

Sobrecarga de rol

Indique su nivel de acuerdo con las siguientes afirmaciones:

1. Tengo demasiado trabajo como para hacerlo todo bien.
2. La cantidad de trabajo que he de realizar (o que me piden que realice) es excesiva.
3. Habitualmente, me falta tiempo para realizar mi trabajo.

Clima de apoyo

Las afirmaciones que le presentamos a continuación hacen referencia a su unidad o grupo de trabajo. Indique su grado de acuerdo o desacuerdo con cada una de ellas.

1. En mi unidad de trabajo la gente se ayuda mutuamente para sacar adelante el trabajo.
2. En mi unidad de trabajo las relaciones interpersonales son buenas.
3. En mi unidad de trabajo las personas muestran interés y apoyo por los problemas personales de los compañeros.