

# **DISEÑO Y VALIDACIÓN DE VARIABLES ESCALARES EN ENCUESTAS: COMPORTAMIENTO DE INDICADORES EMPÍRICOS Y LA PERCEPCIÓN DE LOS ENTREVISTADOS.**

---

Grupo de Trabajo 01. Sesión 1ª

Autor: **ARROYO MENÉNDEZ, Millán.**

Institución de procedencia: UCM

Facultado de Ciencias Políticas y Sociología. UCM

Email: [millan@cps.ucm.es](mailto:millan@cps.ucm.es)

## **1- Planteamiento del problema**

Muchos de los objetos de estudio de encuestas son difíciles de aprehender y tanto más de cuantificar (valores, actitudes, valoraciones o percepciones en general, niveles de satisfacción, etc) cuando se operativizan diseñando indicadores concretos. Estos indicadores son sumamente lábiles, ya que pequeñas variaciones o cambios no sustantivos en la formulación de la pregunta o en las categorías de respuesta, a menudo dan lugar a resultados muy diferentes en términos de distribución de respuestas (Rugg, 1941). Cabe plantearse si no también, a veces, en la cualidad conceptual de lo que se recoge.

Resulta, por ejemplo, suficientemente conocido que la posición de una pregunta en el cuestionario puede cambiar significativamente la distribución de frecuencias obtenidas, así como afecta también la posición de un ítem concreto en una batería de respuestas en las menciones finalmente obtenidas en dicho ítem. (Schuman & Kalton, 1985 p 651; Schumman & Presser, 1981, pp 23-74).

También es sabido que la inclusión u omisión de categorías de respuesta concretas afecta decisivamente a los resultados finales, es decir, a la redistribución final de dichas categorías. Respecto a esto último, se conocen resultados experimentales sobre los efectos de incluir o no alternativas intermedias en preguntas sobre actitudes y asociaciones. (Schumman & Presser, 1981). Así por ejemplo, si queremos que los entrevistados se autclasifiquen utilizando los términos 'liberal' y 'conservador' los resultados obtenidos son sustantivamente diferentes si incluimos o no una categoría intermedia, neutral entre ambos, del tipo 'intermedio' o 'término medio'. También varía la proporción de no respuesta obtenida.

Al problema de recoger la información de distintas maneras, dependiendo del diseño del indicador, se le añade el problema de la 'no cristalización' de las opiniones y actitudes que las preguntas de encuesta tratan de aprehender. Los entrevistados pueden estar expresando opiniones sobre las que anteriormente no han reflexionado y son por tanto poco consistentes. (Schumman & Presser, 1981).

Por todos estos problemas, y otros no mencionados, en no pocas ocasiones de la práctica de encuesta, surgen dudas sobre los indicadores utilizados, cuando en el contexto del análisis el indicador se comporta de forma inesperada, y sobreviene la sospecha de hasta qué punto refleja adecuadamente el concepto que desea medir, o nos planteamos hasta qué punto podemos fiarnos de su 'finura'. El mayor peligro sin embargo es no percatarse de la fragilidad del constructo y tratar de sacar conclusiones sobre los datos de una manera acrítica, sin una debida reflexión sobre el contexto de la recogida de la información, o sin estudiar su comportamiento empírico real, por ejemplo, interaccionando con otras variables.

Con el ánimo de conocer mejor el comportamiento empírico de algunos de los indicadores escalares de encuestas con los que más a menudo se trabaja, en el contexto (semántico) de la sociedad española y el uso de la lengua castellana, se ha realizado una pequeña encuesta en la que se han preguntado reiteradamente las mismas cuestiones, pero recogiénolas de forma diferente. Se han examinado tres grupos de fenómenos: la satisfacción con los estudios, la identificación subjetiva de clase y el precio que estarían dispuestos a pagar por un reproductor de DVD's. La encuesta (de una sola hoja) se realizó mediante cuestionario autoadministrado dirigido a estudiantes universitarios de diversas carreras de ciencias sociales<sup>1</sup> en la Universidad Complutense de Madrid y en la Carlos Tercero de Madrid. (Relaciones laborales, de la UCM, Humanidades, de la UC3M, y Administración y Dirección de empresas de la UC3M). Respondieron finalmente un total de 138 estudiantes<sup>2</sup>. Como complemento a la encuesta se han realizado 11 entrevistas abiertas, a individuos que previamente habían respondido al cuestionario, para relacionar las respuestas de encuesta con el discurso de los sujetos sobre los temas planteados<sup>3</sup>. El cuestionario se presenta anexo al final de este artículo. Las preguntas que se perseguía responder con esta investigación fueron las siguientes:

- *¿Hasta qué punto la presentación gráfica de una escala puede influir en la percepción del fenómeno que tratamos de medir?*
- *¿Son adecuadas las dicotomías para recoger estados de opinión complejos?*

---

<sup>1</sup> Se omitieron estudiantes de sociología, y se cuidó especialmente que no hubiesen recibido hasta el momento de realizarse la entrevista en sus estudios conocimientos sobre diseño de escalas en encuestas.

<sup>2</sup> Trabajo de campo: entre mayo y junio de 2004.

<sup>3</sup> Entrevistas abiertas breves, de unos 10 minutos de duración, realizadas telefónicamente los días 30 de junio y 1 de julio.

- ¿Es mejor incluir o no un punto medio (o neutral) en una escala?
- ¿Es mejor utilizar escalas semánticas o numéricas?

## 2- La influencia de la escala en la percepción de lo que se mide:

Con el propósito de responder a la pregunta de hasta qué punto la presentación de la escala puede modificar la percepción de del fenómeno por el que se pregunta, y como modo de ilustrar con un ejemplo las referidas ‘labilidad’ y ‘fragilidad de los indicadores de encuesta, se me ocurrió la siguiente pregunta: “¿Cuál sería el precio máximo que estarías dispuesto a pagar por un reproductor de DVD’s suponiendo que estuvieses totalmente decidido a comprarte uno?”. E hice responder a dos submuestras aleatorias utilizando escalas de respuesta ayudadas diferentes. Si era cierto que el diseño de la escala influye de algún modo en las respuestas, debería obtener distribuciones de respuestas significativamente diferentes; respuestas de precios más bajos para la escala que explicitaba precios más bajos y viceversa. Las escalas utilizadas fueron las que figuran a continuación y los resultados los que siguen (T1):

<b>ESCALA A</b>	<b>ESCALA B</b>	<b>T1: RESULTADOS</b>	
		<b>ESCALA A</b>	<b>ESCALA B</b>
Menos de 40 €	Menos de 80 €		
De 41 a 60	De 81 a 100		
De 61 a 80	De 101 a 120		
De 81 a 100	De 121 a 140	Menos de 80	53%
De 101 a 120	De 141 a 160	De 81 a 200	44%
De 121 a 140	De 161 a 180	Más de 200	3%
De 141 a 160	De 181 a 200	N=	<b>(64)</b>
De 161 a 180	De 201 a 220		
De 181 a 200	De 221 a 240		
Más de 200	Más de 240		

*Chi cuadrado: 23,895. p<0,004.*

Pese al reducido tamaño de la muestra, las diferencias agregadas en los 3 intervalos son de tal magnitud, que resultan significativas estadísticamente. Aparentemente, los que responden a la escala A están dispuestos a gastar mucho menos dinero que los que responden a la escala B. Sin embargo, no hay motivos para pensar que estas diferencias sean reales, es decir, preexistentes antes de formular la pregunta de encuesta. Las submuestras son muy homogéneas entre sí en edad, sexo, estudios y estilo de vida, y además su segmentación fue aleatoria. ¿Porqué responden entonces distinto a la escala A y a la escala B, teniendo en cuenta que en ambas los límites inferior y superior

quedan abiertos? Un individuo que tenga más o menos claro que su límite está próximo a los 70 euros, puntuaría en la 3ª categoría de la escala A o indistintamente a la 1ª de la B. Y uno que fije su límite en unos 230 euros, respondería a la última categoría de la escala A o indistintamente a la penúltima de la B. Eso sería lo razonablemente esperable si el razonamiento de los sujetos no es sensible al diseño de la escala.

Sin embargo, esto no se produce porque el efecto perceptual de la escala interfiere decisivamente en la respuesta. Tanto más cuanto menor es el conocimiento del sujeto sobre los precios de los productos en el mercado. Según mi experiencia anterior al citado experimento, toda escala de precios, reproducida en una cartulina de respuesta ayudada, crea la sensación de que la distribución presentada refleja (más o menos fielmente) la distribución de precios en el mercado, es decir, que los que se sitúan en el centro son los precios medios, y más allá de los límites superior e inferior lo que puede haber es residual, poco frecuente. Esta información aportada, (deliberada o no, real o falsa) ayuda decisivamente a responder al sujeto. Debido a ello, los individuos a quienes se mostró la escala A recibían la información de un mercado más barato, y respondieron en consecuencia.

### 3- ¿Son adecuadas las dicotomías para recoger estados de opinión complejos?

Una dicotomía o variable discreta no es preferible, en mi opinión, a un diseño escalar cuando la naturaleza del concepto por el que preguntamos es complejo. Un buen ejemplo de ello es preguntar por la satisfacción por algo: satisfacción con el trabajo, con un producto o servicio, con los estudios, con la vida o uno mismo. El ejemplo es además relevante por la frecuencia con la que se pregunta por la satisfacción en las encuestas.

Cuando a la pregunta *¿Podrías definir, por favor, el grado de satisfacción con los estudios que estás realizando?* se plantean respuestas dicotómicas, en términos de ‘satisfecho’ o ‘insatisfecho’, una amplísima mayoría de entrevistados afirman estar ‘satisfechos’. (En nuestra encuesta nada menos que el 88% de los entrevistados están satisfechos con sus estudios universitarios). Muchas empresas, necesitadas de realizar encuestas de satisfacción a sus clientelas para acceder a la certificación de calidad (ISO 9001, 9002 etc) y con poca experiencia en encuestas, plantean el sondeo de la satisfacción en tales términos dicotómicos, lo cual suele llevar a resultados altamente autocomplacientes, ya que los porcentajes de satisfacción que se recogen de esta manera suelen ser altísimos; la insatisfacción se encuentra de este modo muy infraestimada.<sup>4</sup>

Esta forma de recoger la información resulta muy poco fiable, dado que muchos aparentes ‘satisfechos’ no lo están tanto, cuando se profundiza en este asunto con otras preguntas de encuesta de respuesta más matizada o en el contexto de la investigación cualitativa. La insatisfacción parcial no se refleja bien en esta modalidad de respuesta dicotómica, pese a que suele ser el fenómeno más extendido: tanto la satisfacción como su contraria, suelen ser fenómenos graduales, y como tales deberían medirse. La no respuesta (1%) parece obedecer a una situación intermedia, de indefinición entre ambas categorías, que coincide con la expresión ‘ni satisfecho ni insatisfecho’. Esta es al menos la razón (demostrada) por la que en los dos casos de no respuesta, optaron por no responder a la pregunta. En cualquier caso, la tasa de no respuestas en este tipo de situaciones suele ser muy baja, como lo muestra la experiencia demoscópica. Lo habitual resulta encontrar amplísimas mayorías de satisfechos, una minoría de insatisfechos y tasas de no respuesta muy bajas o inexistentes. Como muestra la tabla

---

<sup>4</sup> Mi experiencia como consultor especializado en esta temática así lo avala. Durante 16 años he realizado estudios de satisfacción en: DATA, Emer-GFK, Bureau Veritas y Mapfre Asistencia.

siguiente (T2), la mayoría de los satisfechos no lo están totalmente, sino solo de forma parcial, y no pocos se mueven entre la satisfacción y la insatisfacción, aunque ante la dicotomía se decanten por la primera.

**T2: Grado de satisfacción con los estudios, según distintas formulaciones (P1 y P4)**

Absolutos		P1		Total
		Satisfecho	Insatisfecho	
P4	Totalmente satisfecho	10	0	10
	Bastante satisfecho	59	0	59
	Más bien satisfecho	42	0	42
	Ni satisfecho ni insatisfecho	9	6	15
	Más bien insatisfecho	0	5	5
	Bastante insatisfecho	0	2	2
	Totalmente insatisfecho	0	1	1
	<b>Total</b>	<b>(120)</b>	<b>(14)</b>	<b>(134)</b>

También los insatisfechos manifiestan diferentes grados de satisfacción, cuando se ofrece esta posibilidad de graduar la respuesta. Queda por tanto claro que la satisfacción es un fenómeno que se manifiesta en la realidad de forma gradual, y como tal debe intentar registrarse desde una aproximación cuantitativa. Veamos a continuación los comportamientos de algunos de los indicadores más habituales, utilizados para investigar la satisfacción o fenómenos similares en encuestas.

#### **4- ¿Es mejor incluir o no un punto medio o neutral en la escala?**

Uno de los temas más polémicos a la hora de diseñar una escala, es si incluir o no una categoría intermedia o neutral. Comenzando por las respuestas escalares semánticas, los resultados o conclusiones aparentes a las que se puede llegar son diferentes según optemos por una solución u otra. No obstante, depende de cómo se formulen las categorías de respuesta intermedias, dicha diferencia se puede minimizar.

### T3: Satisfacción con los estudios según formulaciones con y sin punto medio

		P3					
% sobre el total de casos (138)	TOTAL	Total. satisfech.	Bastante satisfech.	Más bien satisfech.	Mas bien insatisfe.	Bastante insatisfe.	Total. Insatisfe.
<b>P4</b>	<b>100</b>	7	45	36	10	1	1
<b>Totalmente satisfecho</b>	<b>7</b>	7					
<b>Bastante satisfecho</b>	<b>44</b>		44				
<b>Más bien satisfecho</b>	<b>30</b>		1	30			
<b>Ni satisfecho ni insatisfecho</b>	<b>12</b>			6	6		
<b>Más bien insatisfecho</b>	<b>4</b>				4		
<b>Bastante insatisfecho</b>	<b>1</b>					1	
<b>Totalmente insatisfecho</b>	<b>1</b>						1
<b>NS/NC</b>							
<b>P2</b>	<b>100</b>	7	45	36	10	1	1
<b>Muy satisfecho</b>	<b>16</b>	7	8	1			
<b>Satisfecho</b>	<b>63</b>		37	26			
<b>Ni satisfecho ni insatisfecho</b>	<b>16</b>			9	7		
<b>Insatisfecho</b>	<b>4</b>				3	1	
<b>Muy insatisfecho</b>	<b>1</b>						1

Por ejemplo, en P2, la alternativa a ‘satisfecho’ o ‘insatisfecho’ es la neutral: ‘ni satisfecho ni insatisfecho’. Esta formulación maximiza la concentración de respuestas en el punto intermedio de la escala. En el caso que nos ocupa no es muy alta dicha concentración (16%) pero en otras situaciones la acumulación de respuestas en el punto medio puede poner en peligro la bondad del indicador, sobre todo cuando sospechamos que bajo la pretendida homogeneidad se oculta la diversidad de matices. Dicha concentración (mayor o menor) se suaviza si antes y después del punto intermedio utilizamos las expresiones ‘más bien satisfecho’ y ‘más bien insatisfecho’. En el caso que nos ocupa, la reducción es pequeña (del 16% pasamos al 12%) pero en otras situaciones la reducción puede llegar a ser mucho mayor. No obstante, el interés de estas expresiones no consiste tanto en su capacidad de reducción del punto medio, como de resultar la estrategia más utilizada, y en mi opinión más interesante, cuando lo que se pretende es su eliminación. Si nos fijamos en la formulación de la P3, en 6 categorías de respuesta sin punto medio, observamos unas distribuciones muy distintas de las que obtenemos en P4 y en P2. Esto se consigue sin forzar demasiado a los entrevistados, como de se deduce al apreciar que no hay individuos que se abstengan de responder. Parece que los individuos que se apuntan al punto medio cuando lo hay, pueden decantarse ‘entre la espada y la pared’ si les ayudamos a hacerlo. Fijémonos en que cuando lo demandaban algunos (no muchos, por cierto) era en la situación de una



dicotomía de términos relativamente extremos. Cuando se suavizan los términos, parece que su referencia es mucho más prescindible. ¿Hasta que punto entonces existe o no un punto medio en las escalas? ¿Es algo artificial, o lo artificial es eliminarlo? La pregunta esconde en realidad una trampa, y uno diría que es irrelevante, pues nos remite a otra más profunda: “¿hasta qué punto han cristalizado los fenómenos de opinión pública de los que las encuestas se ocupan?”, difícil saberlo con exactitud, aunque es probable que no pocas opiniones reflejadas en las encuestas hayan merecido muy poca atención, o ninguna, a muchos entrevistados.

Una situación ‘típica’ en la que encontramos una alta concentración en un punto medio de la escala, es en el caso de la clase social subjetiva. En los países con amplias clases medias, existe una percepción (exagerada) a creer que la situación socioeconómica de uno se corresponde con la situación más general en un país. De modo que la concentración de respuestas en el segmento intermedio de clase suele convertirse en el sesgo típico de este indicador, como es bien sabido por los demoscopistas.

En nuestra encuesta, como se expone a continuación, ante las alternativas de clases ‘alta, media y baja’ nada menos que el 91% de los entrevistados responden que pertenecen a la clase media. Todos somos o nos sentimos de clase media. El problema se suaviza pero no se soluciona cuando las alternativas son: ‘alta, media-alta media-media, media-baja y baja. La mitad de los entrevistados (49%) se siguen apuntando a la clase media-media, debido a la fuerte inercia de creer que estamos justo en el medio de la escala social.

<b>T4: Percepción subjetiva de clase:</b>	
<b>N=(138)</b>	<b>%</b>
<b>P8</b>	100
<b>Alta</b>	4
<b>Media</b>	91
<b>Baja</b>	5
<b>NS/NC</b>	1
<b>P9</b>	100
<b>Alta</b>	1
<b>Media Alta</b>	38
<b>Media media</b>	49
<b>Media baja</b>	12
<b>Baja</b>	1
<b>NS/NC</b>	
<b>P10</b>	100
<b>Alta</b>	1
<b>Media alta</b>	64
<b>Media baja</b>	35
<b>Baja</b>	1
<b>NS/NC</b>	

En esta situación eliminar el punto medio puede parecer una solución razonable, pero dependiendo de cómo se haga, la solución puede traer tantos o más inconvenientes que ventajas. Por ejemplo, (ver T4) en nuestro caso particular, la situación empeora de la P10 a la anterior formulación de la P9, aunque mejora respecto de la P8, a juzgar por la capacidad real de distribuir las respuestas en el público que entrevistamos. (Prácticamente las respuestas se distribuyen en 3 categorías en P9, en 2 en P10 y en 1 en P8).

La cuestión del punto medio o neutral es por tanto difícil de resolver. Parece razonable actuar en función de la experiencia y conocimiento previo de la realidad estudiada.

## 5- ¿Mejor escalas semánticas o numéricas?

Volvamos a ocuparnos de la satisfacción con los estudios. Otra posibilidad de cuantificación consiste en utilizar escalas numéricas en lugar de escalas semánticas. Estas son, comparadas con las anteriores, más abstractas, pero también más intuitivas y ‘libres.’ ¿Da más o menos igual recurrir a una escala semántica o a una escala numérica? En caso negativo, entonces procedería preguntar: ¿funcionan realmente mejor que las escalas semánticas? A continuación se ofrece un caso para introducirnos en esta reflexión?

Para responder a estas preguntas contamos con dos situaciones comparativas distintas; una escala de 5 puntos y otra de 7 puntos, ambas a la vez semánticas y numéricas.

Comenzando con la escala de 5 puntos (ver Tabla 5), en P2 utilizábamos una escala semántica a la pregunta: *¿Y si las categorías de respuesta fuesen las siguientes, cómo describirías tu grado de satisfacción?* (las categorías figuran abajo en la tabla). Mientras que en P7 la formulación era: *“¿Podrías definir tu grado de satisfacción con los estudios que estás realizando utilizando para ello una escala de 1 a 5, en la que ‘1’ es el máximo grado de insatisfacción y ‘5’ el máximo grado de satisfacción?”* (La escala no se representaba gráficamente) Los resultados difirieron más de lo que inicialmente se esperaba.

**T5: Comparación de dos escalas de 5 puntos: numérica (P7) y semántica (P2):**

% sobre el total (n=138)		P2				Total
		Muy satisfecho	Satisfecho	Ni satisfac. ni insatisfec	Insatisfecho	
P7	1,0	1,5%				2,2%
	2,0		1,5%	2,9%	2,9%	7,3%
	2,5		,7%	2,9%		3,6%
	2,7			,7%		,7%
	2,8			,7%		,7%
	3,0	1,5%	31,4%	8,8%	1,5%	43,1%
	3,5		3,6%			3,6%
	4,0	7,3%	23,4%			30,7%
	4,5		,7%			,7%
	5,0	5,8%	1,5%			7,3%
<b>Total</b>		<b>16,1%</b>	<b>62,8%</b>	<b>16,1%</b>	<b>4,4%</b>	<b>100,0%</b>

En primer lugar se observa que en P2 (la escala numérica) un 9,3% de entrevistados dan puntuaciones intermedias, pese que dicha posibilidad en principio no se contemplaba. En otros cuestionarios en los que se utiliza esta escala de 1 a 5 no se suele producir un porcentaje tan elevado de respuestas atípicas o no previstas, aunque ocasionalmente me consta que se producen. (Según mi experiencia suelen ser ocasionales y durante el proceso de supervisión y depuración del cuestionario se eliminan. A menudo el técnico a cargo de la encuesta ni siquiera se percata de su existencia). Si en esta ocasión han llegado a ser tan altas obedece a dos razones; por un lado, porque no se ha representado gráficamente la escala (se sentían más libres que en otras situaciones análogas), y por otro lado porque eran conocedores de que el objetivo del cuestionario era observar el comportamiento de distintos indicadores de satisfacción, y el empeño en ser precisos en sus respuestas ha debido afectar a la emergencia de decimales.

En las entrevistas abiertas, cualitativas, los entrevistados argumentaron que hicieron esto debido a que esta distribución no les permitía expresar con precisión lo que pensaban, ya que se quedaban a caballo de dos números enteros; estaban entre el 2 y el 3 y entre el 3 y el 4. Estos entrevistados percibían una gran distancia entre el 3 y sus inmediatos el 2 y el 4 a la vez que sentía que sus respuestas quedaban comprendidas a medio camino de estos valores. Sin embargo, curiosamente, en la P2, también de 5 puntos, no se recoge ningún comentario sobre que la distancia entre el punto medio y categorías inmediatas sea grande. Parece por tanto que la percepción de ambos indicadores es distinta.

En segundo lugar, y lo que es más esclarecedor, si nos fijamos en la tabla, observamos que la correspondencia gradual entre valores no es nada exacta. Es decir, debíamos esperar, si la percepción es igual o muy semejante, que hubiese una correspondencia absoluta o casi entre los siguientes valores: 1 (máximo grado de insatisfacción), con 'muy insatisfecho', 2 con 'insatisfecho', 3 con 'ni satisfecho ni insatisfecho', 4 con 'satisfecho' y 5 con 'muy satisfecho'. Las respuestas son muy distintas, incluso abundan las contradicciones. Por ejemplo: la mayor parte de insatisfechos dan puntuaciones de 4 y no de 5, e incluso puntúan 3 ó 1 (máxima insatisfacción). La mayoría de los 'satisfechos' puntúan 3 y así lo hacen algunos insatisfechos. ¿Qué quiere decir todo esto?.

La P2 (escala semántica) es coherente con otras escalas semánticas (P1, P3 y P4) si atendemos a los cruces realizados mediante tablas de contingencia. Pero la P7 no es coherente ni con P2 ni con ninguna otra pregunta sobre el nivel de satisfacción con los estudios (ateniéndonos al mismo criterio). Tampoco es coherente con P5, otra pregunta de escala numérica sobre satisfacción con los estudios. Cuando preguntamos abiertamente a los sujetos, en un contexto cualitativo, nos encontramos con la principal razón de estas diferencias. Los entrevistados a menudo se confunden con esta escala y cometen errores de apreciación, porque tienen dificultades para reproducir en su mente las dimensiones de la escala. Como consecuencia, a veces no se tiene claro que el punto medio sea el 3, o a veces invierten erróneamente su sentido. Algunos ejemplos ilustran los problemas que plantea la escala a los entrevistados:

Una estudiante de humanidades de la UC3M, de 22 años, afirmaba estar *'bastante satisfecha'* en P3 y en P4, en P5 (escala de 0 a 10) otorga un 8 a su grado de satisfacción y un 6 en P6 (escala de 1 a 7). Al llegar a P7 (escala de 1 a 5) su puntuación es un 3,5. Su argumentación durante la entrevista abierta es consonante con todas estas puntuaciones excepto con la dada en P7, donde cabía esperar un 4. (Estudiaba humanidades porque le gustaba, y estaba bastante contenta con dichos estudios porque habían sido elegidos libremente. Sin embargo su satisfacción no era absoluta porque algunas asignaturas le habían decepcionado, aunque el saldo general era bueno). Al recordarle yo que en P7 su puntuación era de un 3,5 y pedirle explicaciones por tal puntuación, responde sin dudar: *"bueno, sería más bien un cuatro ( ) cuando puse el 3,5 a lo mejor pensé que el 4 era una puntuación muy alta"*. Acto seguido le recordé que el 3 era el punto medio de la escala y respondió: *"creo que no lo tuve en cuenta. Desde luego está mejor el 4"*.

Un estudiante de Administración y dirección de empresas de la UC3M, de 21 años, afirma estar *'más bien satisfecho'* en P3 y en P4. En P5 (escala de 0 a 10) puntúa con un 7. Su argumentación cualitativa es congruente con estas puntuaciones, y además se ratifica en ellas. Se siente satisfecho, aunque no lo suficiente para afirmar que *'bastante'* porque está molesto con que le hayan obligado a asistir a clase por la tarde. En P7 puntuaba con un 3. Al pedirle explicaciones de dicha puntuación respondió: *"Poner un 3 creo que no es correcto, pero tampoco sería un 4. Un 4 es demasiado. No sé qué poner. Creo que la escala se queda corta"*. Sin embargo, en P2 (la semántica de 5

puntos) se otorga un 'satisfecho' (el equivalente teórico de un 4) y se ratifica. En este caso, se ha percibido un gran distancia entre el 3 y el 4, y al rellenar el cuestionario no se percibió el 3 como punto medio. Solo se percibe así después de que yo lo comento.

La representación gráfica de dichas escala, hubiese ayudado en gran medida a paliar estas confusiones, pero hubiese seguido siendo una escala más abstracta que su simétrica semántica, y la percepción y sus resultados, probablemente hubiesen seguido siendo diferentes.

Para averiguar hasta qué punto la representación gráfica de la escala ayuda a evitar errores a los entrevistados, hemos incluido en nuestro cuestionario dos modalidades de la pregunta P6 que mide la satisfacción con los estudios en una escala de 7 puntos. La mitad de los cuestionarios distribuidos representaban gráficamente la escala (en adelante P6 tipo A), mientras que la otra mitad no (en adelante P6 tipo B), reproduciendo una situación semejante a la descrita de P5 (la escala numérica de 5 puntos). Los resultados obtenidos fueron verdaderamente interesantes.

En primer lugar, destacamos que siguieron apareciendo decimales en la formulación tipo A (sin escala) mientras que no aparecieron en la tipo B. Puede por tanto afirmarse que la representación gráfica de la escala cuando menos mitiga en gran medida la aparición de respuestas expresadas con números decimales. A pesar de ello, la frecuencia de aparición de decimales es mucho menor en esta escala de 7 puntos que en la escala de 5, y entiendo que es debido a que el rango de respuestas es más amplio. (3 casos de 65 individuos que respondieron al Tipo A, 4,6%. Estas incidencias se reducen prácticamente a la mitad).

En segundo lugar, si comparamos los resultados de ambas modalidades de P6 con su escala semántica simétrica, (de 7 puntos) la P4, observamos que la relación la P6 tipo B (con escala) con la P4 es mayor que la observada entre P4 y P6 tipo A. Lo que se reducen sobre todo son las principales incoherencias, lo que hace suponer que la representación gráfica de la escala reduce las probabilidades de error. Las entrevistas abiertas confirman este supuesto. También en el caso de la P6 se producen errores de interpretación debido a que la escala no queda clara. Se producen como en P5

confusiones debido a que no queda claro si el punto medio es el 3 ó el 4 ó el 5, y a veces hay confusiones sobre el sentido de la escala.

En la formulación tipo B estas confusiones se reducen bastante, pero no llegan a desaparecer del todo. Entendemos que el contexto de entrevistas auto-rellenadas (como esta encuesta) favorecen este tipo de errores, mientras que la presencia de un entrevistador bien entrenado y motivado puede contribuir a una reducción mucho mayor.

Algunos ejemplos que ilustran los problemas encontrados con la P6 son los siguientes:

Una estudiante de Administración y Dirección de empresas, de la UC3M, de 20 años, afirma estar más bien satisfecha y así se ratifica en la entrevista cualitativa. En P6 (tipo B) puntúa con un 3. Al pedir explicaciones enseguida reconoce que invirtió el sentido de la escala: *“Es un error, debería ser un 5”*. Como vemos, la representación gráfica de la escala no elimina este tipo de errores de apreciación en las escalas de 1 a 5 y de 1 a 7.

Un estudiante de Relaciones Laborales de la UCM puntúa 4 en P6 (tipo B) y 3 en P7 (la escala de 5 puntos). Estas dos puntuaciones intermedias contradicen su discurso cualitativo y el resto de las puntuaciones dadas. Al pedir explicaciones sobre la razón de estas puntuaciones, tras advertir yo que había elegido los puntos intermedios responde: *“[En P6] he entendido el 4 como algo más que neutral, más por arriba que lo neutral [recordamos que la escala se representaba]. [En P7] no era mi intención elegir la puntuación intermedia, sino algo más, algo más tirando hacia arriba”*. Dicho sea de paso, muchos de los que tuvieron problemas en P7 con la escala de 1 a 5, también tuvieron problemas en P6, con la escala de 1 a 7. Las confusiones eran del mismo tipo, aunque se agudizaban en P7.

Una estudiante de Administración y dirección de empresas de la UC3M se considera más bien insatisfecha. Sin embargo en P6 (tipo A) otorga un 4 (puntuación intermedia). Al pedirle explicaciones y hacerla ver que había dado una puntuación intermedia, responde: *“Pensaba que el 4 era más bajo. Estoy un poco insatisfecha. Se supone que debería haber puesto un 2 ó un 3”*.

Lo que quedaba claro en las entrevistas cualitativas, es que los entrevistados se ratificaban en las respuestas dadas en las escalas semánticas, y estas eran coherentes con las explicaciones abiertas sobre la satisfacción con los estudios. Lo que a veces entraba en contradicción con sus discursos eran las respuestas dadas en las escalas numéricas, y al pedir explicaciones, reconocían el error de interpretación en las escalas numéricas y redefinían sus puntuaciones resultando entonces coherentes con sus discursos y otras respuestas de encuesta.

Ante estos restos resultados, creo que puede responderse a las preguntas iniciales, afirmando que las escalas semánticas ofrecen resultados diferentes que las escalas numéricas, que el principal problema de estas últimas es que presentan problemas de asimilación y retención por parte del entrevistado, lo que redundaba en confusiones e imprecisiones en las respuestas. Este problema se da sobre todo en las escalas que no se representan gráficamente, y se reducen bastante aunque no desaparecen en las que sí se representan. Al margen de lo dicho, y con independencia de ello, su carácter más abstracto hace que la percepción de estas escalas sea diferente al de sus simétricas semánticas.

**T6 Relación entre escalas de 7 puntos, P4 semántica y P6 numérica (p6a escala no dibujada y p6b dibujada)**

TIPO	Absolutos		P4					Total	
			Total. satisfecho	Bastante satisfecho	Más bien satisfecho	Ni satisf ni insatisfec	Más bien insatisfecho		Bastante insatisfe.
<b>A</b>	<b>P6</b>	1	1	0	0	0	0	0	1
		3	0	0	1	2	0	0	3
		3,5	0	0	1	1	0	0	2
		4	0	3	5	4	3	0	15
		5	1	16	11	0	0	0	28
		5,5	0	1	0	0	0	0	1
		6	2	9	0	0	0	0	11
		7	4	0	0	0	0	0	4
		Total	8	29	18	7	3	0	65
<b>B</b>	<b>P6</b>	1	0	0	0	0	0	0	1
		2	0	1	1	0	1	2	5
		3	0	0	3	3	2	0	8
		4	0	1	10	6	0	0	17
		5	0	14	7	0	0	0	21
		6	1	16	2	0	0	0	19
		7	1	0	1	0	0	0	2
		Total	2	32	24	9	3	2	73



**T8: Escala de 0 a 10 comparada con escala semántica:**

		P4							
Absolutos		Totalmente satisfecho	Bastante satisfecho	Más bien satisfecho	Ni sat. ni insatisfecho	Más bien insatisfecho	Bastante insatisfecho	Totalmente insatisfecho	Total
<b>P5</b>	0	0	0	0	0	0	0	1	1
	1	1	0	0	0	0	1	0	2
	3	0	0	1	0	1	1	0	3
	4	0	0	0	0	3	0	0	3
	5	0	0	8	14	2	0	0	24
	5,5	0	0	1	0	0	0	0	1
	6	0	1	14	2	0	0	0	17
	6,5	0	2	0	0	0	0	0	2
	7	1	24	16	0	0	0	0	41
	8	0	27	2	0	0	0	0	29
	9	3	6	0	0	0	0	0	9
	10	5	0	0	0	0	0	0	5
Total		10	60	42	16	6	2	1	137

Con todo, no todas las escalas numéricas se encuentran en situación de desventaja frente a las semánticas. En mi opinión, y experiencia, las escalas de 0 a 10 ó de 1 a 10 (indistintamente) ofrecen comportamientos mucho más ventajosos y satisfactorios que los de las escalas anteriores, y son especialmente recomendables para cuantificar valoraciones, tanto o más que las escalas semánticas. Tienen la ventaja de un rango amplio de graduación de la respuesta, y carecen del inconveniente de la abstracción e inconcreción de los valores, ya que es una escala que se conoce muy bien, por haber sido utilizada en el sistema educativo para calificar y evaluar. Todo el mundo sabe que el 5 es ‘un aprobado’, un 7 ya es un ‘notable’, un 4 es un ‘suspense’, etc. Estos baremos están tan interiorizados, que no es preciso recordarlos. Además, la representación gráfica no es necesaria por el mismo motivo.

En esta encuesta, la escala de 0 a 10 (P5) mantiene una relación mucho más estrecha y menos contradictoria con las escalas semánticas. Su congruencia se ha podido ver en el contexto de las entrevistas cualitativas, analizando los diversos casos particulares. Desde una perspectiva cuantitativa también apreciamos esta congruencia. Si la cruzamos por ejemplo con la P4 (la semántica más larga, con 7 intervalos) observamos que las incongruencias prácticamente desaparecen (sólo hay 2 casos realmente incongruentes, de un total de 138, equivalente al 1,4%, algo asumible trabajando con encuestas) y una relación de categorías de respuesta que es la que sigue: los totalmente satisfechos, suelen dar las puntuaciones extremas de 9 y 10. Los bastante satisfechos se

corresponden con las puntuaciones de 7 y 8. Los más bien satisfechos oscilan entre el 5 y el 7. Los ni satisfechos ni insatisfechos en el 5. Los bastante insatisfechos entre el 1 y el 3 y los totalmente insatisfechos en el 0.

## **6- Sugerencias para diseñar y validar indicadores escalares de encuestas:**

De todo lo dicho hasta ahora se desprenden dos conclusiones que desearía destacar. La primera, que el ‘fetichismo’ del dato es un absurdo, dado que los resultados dependen estrechamente de la formulación de la pregunta y de las categorías de respuesta ofrecidas. No hay por tanto precisión posible en los indicadores de encuestas si esta se desliga del contexto de formulación (y por tanto, de la herramienta de medición), al menos en fenómenos tales como actitudes, valores o niveles de satisfacción.

La segunda, que el trabajo de diseño de un indicador se revela como una tarea compleja y delicada a la que habría que dedicar una atención especial, dado que la forma de preguntar no da en absoluto igual; ya hemos visto que influye decisivamente. Las personas poco acostumbradas a trabajar con encuestas, suelen ser ‘fetichistas del dato’ y cuando descubren (o se les explica) la variación de los resultados se desconciertan tremendamente y pasan radicalmente de la fe del carbonero al ateísmo en materia de encuestas. Que la labor sea delicada no significa la invalidez de la herramienta. Al contrario, la herramienta es válida el propósito de la investigación social, siempre y cuando tomemos ciertas precauciones.

En primer lugar, debe valorarse el conocimiento previo del investigador a la hora de plantear un cuestionario, tanto en lo que se refiere a la materia objeto de estudio como a su experiencia recogiendo y analizando información a través de encuestas. La experiencia y la intuición del investigador cuantitativo pueden y deben jugar un papel tan relevante o reconocido como en la investigación cualitativa, pues su actividad no es ni más ‘científica’ ni menos ‘artesanal’ que la anterior.

En segundo lugar, el análisis de la información no debe ser ingenuo, es decir, debe hacerse consciente de la dependencia de la forma de preguntar sobre los resultados, lo

cual plantea la necesidad de validar el indicador construido antes de extraer conclusiones sobre los datos obtenidos.

Mi experiencia profesional, de 16 años a cargo de estudios demoscópicos, me hace ver que en ambas situaciones de diseño y validación de indicadores escalares, la investigación cualitativa resulta de gran ayuda, y supone un recurso obligado cuando uno desea evitar errores de apreciación o interpretación. Esta se puede incorporar al trabajo con indicadores de encuesta de varias maneras:

- De forma exploratoria, incorporando una fase cualitativa previa al diseño de una encuesta, donde la formulación de las preguntas se intente aproximar a los argumentos discursivos y a las expresiones de los informantes, y donde se extraiga información para diseñar las categorías de respuestas.
- Pilotando cualitativamente las preguntas y sentido dado a las categorías de respuestas, una vez diseñado el cuestionario y antes de lanzarse a campo.
- Ampliando la información y resolviendo las incógnitas que a posteriori plantea la explotación cuantitativa de la encuesta. Es muy interesante y enriquecedor seleccionar perfiles tipológicos elaborados a partir de preguntas de encuesta y contrastar sus actitudes y opiniones cuantitativas con su discurso recogido cualitativamente, y la observación del contexto en que dicho discurso se teje.

Aparte del recurso a la exploración cualitativa, a priori e incluso a posteriori, me parece imprescindible para validar los indicadores creados, el contraste cuantitativo. Consiste en formular hipótesis lógicas, más o menos esperables e idealmente ya contrastadas o sabidas, sobre la relación del concepto que se mide a través de un indicador con otros conceptos o fenómenos sociales, de los que tenemos información en el mismo cuestionario. Si el indicador empírico se comporta conforme lo esperado, entonces cabe concluir que es razonablemente bueno, mientras que si su comportamiento es anómalo cabe plantearse que no responde del todo a lo que pretendíamos y por tanto no es tan bueno.

Si recogemos de más de una manera un mismo concepto, esta prueba de consistencia teórica puede servir para decidir cual es mejor de los dos.

## BIBLIOGRAFÍA:

---

CONVERSE, J.M. & PRESSER, S. 1988. *Survey questions. Handcrafting the standard questionnaire*. SAGE University Paper. N° 63. Beverly Hills, California.

LODGE, M. 1981. *Magnitude Scaling. Quantitative Measurement of Opinions*. SAGE University Paper. N° 25. Beverly Hills, California.

RUGG, D. (1941) "Experiments in wording questions: II". En: *Public. Opinion. Quart...*, 5, 91-92.

SCHUMAN, H. & KALTON, G. 1985. "Survey Methods" (Chapter 12, pp 635-667) en: *The Handbook of Social Psychology*. Ed. Gardner Lindzey & Elliot Aronson. Random House, New York.

SCHUMAN, H. & PRESSER, S. 1981. *Questions and answers in attitude surveys. Experiments on question form, wording, and context*. Academic Press, INC. London.

**INTRODUCCIÓN:** Esta breve encuesta responde a una investigación sobre el comportamiento de escalas de respuestas que lleva a cabo el profesor Dr. Millán Arroyo Menéndez. Para responder circula el código de respuesta que se corresponde en cada pregunta a la respuesta elegida. Intenta no dejar en blanco ningún dato.

P1 ¿Podrías definir, por favor, el grado de satisfacción con los estudios que estás realizando?

Satisfecho 1  
 Insatisfecho 2

P2 ¿Y si las categorías de respuesta fuesen las siguientes, cómo describirías tu grado de satisfacción?

Muy satisfecho 1  
 Satisfecho 2  
 Ni satisfecho ni insatisfecho 3  
 Insatisfecho 4  
 Muy Insatisfecho 5

P3 ¿Y si fuesen estas?

Totalmente satisfecho 1  
 Bastante satisfecho 2  
 Más bien satisfecho 3  
 Más bien insatisfecho 4  
 Bastante insatisfecho 5  
 Totalmente insatisfecho 6

P4 ¿Y si fuesen estas?

Totalmente satisfecho 1  
 Bastante satisfecho 2  
 Más bien satisfecho 3  
 Ni satisfecho ni insatisfecho 4  
 Más bien insatisfecho 5  
 Bastante insatisfecho 6  
 Totalmente insatisfecho 7

P5 ¿Podrías definir tu grado de satisfacción con los estudios que estás realizando utilizando para ello una escala de 0 a 10, en la que '0' es el máximo grado de insatisfacción y '10' el máximo grado de satisfacción?

\_\_\_\_\_

P6 ¿Podrías definir tu grado de satisfacción con los estudios que estás realizando utilizando para ello una escala de 1 a 7, en la que '1' es el máximo grado de insatisfacción y '7' el máximo grado de satisfacción?<sup>5</sup>

1 2 3 4 5 6 7

P7 ¿Podrías definir tu grado de satisfacción con los estudios que estás realizando utilizando para ello una escala de 1 a 5, en la que '1' es el máximo grado de insatisfacción y '5' el máximo grado de satisfacción?\_\_\_\_\_

P8 ¿Podrías indicar, de acuerdo con las siguientes categorías de respuestas, cual es la clase social a la que perteneces tú y tu familia?

Alta 1  
 Media 2  
 Baja 3

P9 ¿Cómo definiríais a ahora la clase social tuya y de tu familia si las categorías son estas:

Alta 1  
 Media alta 2  
 Media media 3  
 Media baja 4  
 Baja 5

P10 ¿Y si las categorías fuesen estas, cual será la ubicación más aproximada de tu clase social?

Alta 1  
 Media alta 2  
 Media baja 3  
 Baja 4

<sup>5</sup> Versión B de esta pregunta. La versión A no incluía la representación gráfica de la escala

P11 ¿Cuál sería el precio máximo que estarías dispuesto/a a pagar por un reproductor de DVD's, suponiendo que estuvieses totalmente decidido a comprarte uno?<sup>6</sup>

Menos de 40 euros	1	Menos de 80 euros
De 41 a 60	2	De 81 a 100
De 61 a 80	3	De 101 a 120
De 81 a 100	4	De 121 a 140
De 101 a 120	5	De 141 a 160
De 121 a 140	6	De 161 a 180
De 141 a 160	7	De 181 a 200
De 161 a 180	8	De 201 a 220
De 181 a 200	9	De 221 a 240
Mas de 200	10	Mas de 240

P12 Respecto a los estudios, ¿cuál es tu grado de satisfacción en relación con estos aspectos, siguiendo esta escala: muy satisfecho, bastante satisfecho, poco satisfecho, o nada satisfecho?

	<u>MUY</u> <u>SATISFECHO</u>	<u>BASTANTE</u> <u>SATISFECHO</u>	<u>POCO</u> <u>SATISFECHO</u>	<u>NADA</u> <u>SATISFECHO</u>
• Compañeros .....	4	3	2	1
• Profesores .....	4	3	2	1
• Como capacitación para el trabajo (utilidad práctica) .....	4	3	2	1
• Métodos de enseñanza.....	4	3	2	1
• Organización del Centro: normas, participación .....	4	3	2	1
Satisfacción con los estudios en general	4	3	2	1

Sexo: Varón 1 Mujer 2

Edad (anotar edad exacta) \_\_\_\_\_

Estudios: \_\_\_\_\_

Curso: (indicar nº de curso) \_\_\_\_\_

*El tratamiento de los datos es totalmente anónimo. No obstante, a algunos entrevistados se les va a requerir para que, en una entrevista abierta, den explicaciones más detalladas de lo que han querido decir al responder a algunas de las preguntas de esta encuesta. Si no te importa ser citado para ser de nuevo entrevistado/a, por favor rellena estos datos. Si tienes algún reparo, puedes dejarlos en blanco:*

Nombre y Apellidos: \_\_\_\_\_

Forma de contacto: (teléfono y/o email) \_\_\_\_\_

**MUCHAS GRACIAS POR TU COLABORACIÓN**

<sup>6</sup> Las categorías de la columna de la izquierda se distribuyeron aleatoriamente en el 50% de los cuestionarios. Y las de la derecha se distribuyeron en el otro 50% también aleatoriamente.