



Efectos del cambio de orden de variables y categorías de respuesta en los resultados de una encuesta política

Perez-Sindin, Xaquín; Abella Vázquez, Carlos Manuel ; Rodríguez Losada, Diana

Published in:
Metodología de Encuestas

Publication date:
2010

Document version
Også kaldet Forlagets PDF

Document license:
[Ikke-specificeret](#)

Citation for published version (APA):
Perez-Sindin, X., Abella Vázquez, C. M., & Rodríguez Losada, D. (2010). Efectos del cambio de orden de variables y categorías de respuesta en los resultados de una encuesta política. *Metodología de Encuestas*, 12(12), 7-27.

EFFECTOS DEL CAMBIO DE ORDEN DE VARIABLES Y CATEGORÍAS DE RESPUESTA EN LOS RESULTADOS DE UNA ENCUESTA POLÍTICA

**Carlos Manuel Abella Vázquez, Diana Rodríguez Losada y
Xaquín Pérez-Sindín**

Avanza Técnicas de Investigación S.L.

C/ Doctor Fleming 78+80 Bajo

15401 Ferrol (A Coruña)

Tfno. 981 946 347

carlos@analzatis.com

RESUMEN: La correcta definición de un cuestionario es una parte esencial del proceso de investigación mediante encuestas para la obtención de información fiable, válida y de calidad. Por ello, en su diseño deben cuidarse al máximo aspectos como la forma y el orden en que se formulan las cuestiones, las opciones de contestación ofrecidas o la redacción tanto de las preguntas como de las categorías de respuesta. En el presente artículo se exponen los resultados de un experimento consistente en administrar dos cuestionarios similares sobre temas políticos a dos muestras extraídas de un mismo universo (*split-ballot experiment*). Las variaciones entre ambos cuestionarios son pequeñas y aluden a cambios de redacción de preguntas para medir un mismo concepto, a variaciones en el lugar que las variables ocupan en el cuestionario y cambios en el orden de presentación de las categorías de respuesta. Los resultados muestran que determinadas variables han sido especialmente sensibles a su cambio de lugar en el cuestionario y que el orden de presentación de las categorías de respuesta puede influir significativamente en los resultados, mientras que el cambio de redacción apenas influyó.

PALABRAS CLAVE: *split ballot experiment*, efectos redacción cuestionarios, orden preguntas, orden respuestas, sesgo.

ABSTRACT: One of the most important issues regarding survey research is the design of questionnaires. Its correct definition constitutes an essential point in the process of obtaining valid, reliable and quality data. For this reason, it is necessary to pay full attention to aspects such as wording or questions and answers order. A split-ballot experiment was conducted to test the hypotheses that slight changes in wording, question placement on the questionnaire and reading of answers could have in the results of a political survey. The results show that some variables were especially sensitive to a variation of placement on the questionnaire and changes in the order in which the answers were read. At the same time, a different wording of the same concept did not show significant differences.

KEYWORDS: split ballot experiment, wording questionnaires effects, questions order, answers order, bias.

Recibido: 27 de noviembre 2008

Revisado: 8 de enero 2009

Aceptado: 17 de febrero 2009

Introducción

El presente estudio se basa en la utilización de la técnica *split ballot*, esto es, la división de una muestra en dos mitades y la aplicación de un cuestionario ligeramente diferente en cada una de ellas. Se pretende comprobar si dichas modificaciones pueden variar los resultados finales o no. Más concretamente, se han querido testar tres tipos de cambios:

- Cambios en el orden de lectura de las preguntas.
- Cambios en el orden de presentación de las categorías de respuesta.
- Variaciones en la redacción de la pregunta y en las categorías de respuesta, manteniendo equiparable el concepto que se pretendía medir.

Una vez diseñados ambos cuestionarios, uno ha sido pasado al Grupo de Control (GC en adelante) y el otro al Grupo Experimental (GE en adelante). Las preguntas más “ortodoxas”, esto es, aquellas redactadas conforme a los usos comunes de los cuestionarios más estandarizados, han sido incluidas en el GC. Entran en esta categoría, por ejemplo, las variables que emplea habitualmente el CIS.

La presentación de los resultados de este experimento se estructura en cuatro apartados, tal y como se detalla a continuación.

En el punto dos se presentan el marco teórico y las hipótesis de partida que estructuran el estudio en su conjunto. Aquí se explica detenidamente qué se pretendía testar con cada modificación introducida en el GE y cuáles eran sus fundamentos teórico-metodológicos.

En el tercer apartado se detalla la metodología seguida para llevar a cabo el experimento. Se especifica también aquí la ficha técnica del estudio.

En el cuarto apartado se realiza la exposición de los resultados del estudio, mientras que en el quinto se sintetizan las principales conclusiones del experimento.

Marco teórico e hipótesis de partida

a) Fundamentos teóricos

El cuestionario supone el pilar fundamental de toda investigación mediante encuestas, puesto que es el medio a través del cual se recogen y estandarizan las respuestas de las personas entrevistadas, lo que nos proporciona los datos necesarios para efectuar posteriormente el análisis estadístico deseado.

A pesar de que de forma genérica se reconoce la importancia de esta fase, a menudo los esfuerzos de cualquier equipo investigador se dirigen a los siguientes pasos y, especialmente, al diseño muestral. Y, en relación con esta etapa, sí se puede afirmar que a lo largo de los últimos años se ha conseguido un elevado grado de sofisticación y refinamiento que permite la ejecución de complejos trabajos de campo con unas elevadas garantías técnicas.

Tal grado de desarrollo y refinamiento no se ha trasladado con igual diligencia, sin embargo, a la propia construcción de los cuestionarios. En esta etapa todavía

prevalece, en cierto modo, la idea de que cualquier cuestionario es igualmente válido mientras respete una serie de reglas genéricas más o menos aceptadas (Krosnick, 1999:538).

En este sentido, es preciso afirmar que ningún diseño muestral posterior, por complejo y refinado que sea, podrá paliar el déficit de calidad de los datos. Y que, en consecuencia, ningún análisis estadístico ulterior, por sofisticado y avanzado que sea, compensará este fallo inicial. No está de más recordar el acrónimo metodológico que sintetiza esta idea: GIGO (Garbage In – Garbage Out). Como convenientemente recuerdan Glendall y Hoek, “the quality of data analysis depends on the quality of the data analysed” (2002).

Por este motivo, averiguar si existen cambios significativos en los patrones de respuesta a un cuestionario si se le aplican determinadas variaciones menores es un tema de especial interés en el campo de los estudios cuantitativos centrados en las encuestas, al afectar a los fundamentos teórico-metodológicos de las mismas.

Hace ya más 60 años, uno de los fundadores de la moderna investigación cuantitativa afirmaba que los resultados de las encuestas estaban más influidos por el diseño de las preguntas que por el diseño muestral (Gallup, 1947, citado en Glendall y Hoek, 2002). Más recientemente, Krosnick (1999) recordaba que mientras los métodos de muestreo habían alcanzado algo así como una práctica estándar, el diseño de cuestionarios todavía era visto más como un arte que como una ciencia.

Desde entonces, se han llevado a cabo múltiples investigaciones analizando los efectos que el diseño de preguntas y respuestas tiene en los resultados finales, la mayor de ellas parte fuera de nuestro país. A continuación se realiza un breve recorrido histórico por estos experimentos.

Glendall y Hoek (2002) analizaron los efectos de la redacción de las preguntas en los resultados finales. La metodología utilizada en dicho estudio se basó en un *split ballot experiment*. Se condujo una encuesta presencial a una muestra dividida, cada mitad compuesta por 347 y 311 unidades. El estudio del que partieron es un ómnibus anual del Departamento de Marketing de la Universidad de Massey, realizado en los hogares de Palmerson North.

Entre las conclusiones de su estudio, dos merecen ser destacadas. En primer lugar, que dichos efectos son potencialmente importantes, pero a menudo impredecibles. En segundo lugar, que es posible formular una misma pregunta (esto es, medir un mismo concepto) de diversas formas sin que por ello se vean especialmente afectados los resultados finales¹. En el presente estudio hemos hecho este experimento con el índice de popularidad, como se verá en el apartado .

La variación del orden de las preguntas también ha sido objeto de estudio. Así, por ejemplo, McAllister y Wattengerg (1995) se plantearon si los niveles de identificación partidista podían variar en función del orden con que se preguntara por la intención de voto y la simpatía política². Con datos de Estados Unidos y Gran

¹ Para una revisión de diversos estudios centrados en este tema, véase Krosnick, 1989.

² Holbrook, Krosnick y Moore (2005) analizaron los efectos del orden las respuestas en preguntas con categorías dicotómicas. Carlson (2005) estudió el impacto del orden de las

Bretaña, concluyeron que el efecto era nulo en el primer caso y muy pequeño en el segundo. Las encuestas de partida fueron la Encuesta Electoral Británica de 1992, que también incluía un experimento de muestra dividida, y el Estudio del Panel Electoral Americano. La idea de alterar el orden de presentación de ambas variables ha sido replicada en el presente estudio, y los resultados se presentan más adelante.

Tourangeau, Singer y Presser (2003) testaron la influencia que la alteración del orden de las preguntas tenía en los resultados de las respuestas en una investigación sobre actitudes. El estudio se basó en encuestas telefónicas llevadas a cabo por Gallup entre julio y octubre de 1999 y entre abril y julio de 2000 con muestras de 1.677 y 1.978 unidades respectivamente. Los resultados obtenidos fueron realmente interesantes. En cuatro de los cinco experimentos realizados se hallaron diferencias significativas con el cambio de orden. No obstante, se demostró asimismo que la alteración del orden no influía en las preguntas posteriores del cuestionario, es decir, dicho cambio tenía un efecto local.

Más recientemente, en un breve pero sugerente artículo ilustrativamente subtítulo "Change The Order Of The Questions In A Poll, You Change The Outcome", la directora de encuestas de la CBS Kathy Frankovic (2008) exploraba la influencia que el orden de presentación de determinados políticos norteamericanos podía tener en sus valoraciones e índices de popularidad. La metodología para este estudio partía del análisis de dos encuestas realizadas por la CBS con un mes de diferencia. En este caso, por tanto, no se optó por un *split ballot experiment*, sino por la comparación de datos diacrónicos, con la posible incidencia que esto pudiese tener en los resultados, máxime teniendo en cuenta la importancia que la coyuntura sociopolítica tiene en la valoración de estos temas. No obstante, la idea ha sido retomada en el presente estudio y trasladada al contexto español, como se verá más adelante.

En España, una experiencia pionera sobre este tema fue llevada a cabo por Díaz de Rada (2002), basándose en datos de una investigación sobre hábitos de consumo realizada en Navarra en 1998. Este estudio, realizado a través de encuestas presenciales, demostraba un efecto primacía en las preguntas analizadas, tanto si se presentaba tarjeta como si se leían las respuestas sin esta ayuda visual. Este último resultado contrasta con investigaciones realizadas en otros contextos, que dan un efecto de "última opción" en este tipo de preguntas. Este mismo efecto de primacía se repite en las preguntas que se responden con escalas, si bien las diferencias no se encuentran en los extremos, sino en las categorías más suavizadas ("de acuerdo" frente a "en desacuerdo"). Finalmente, este estudio concluye que apenas existen diferencias entre estos efectos cuando se analiza en detalle la influencia del nivel educativo y de la edad del entrevistado.

En definitiva, el objetivo del presente estudio es obtener una visión panorámica y fundamentalmente descriptiva sobre los cambios que se pueden producir en los resultados finales si se alteran determinadas preguntas. El instrumento de medida ha

preguntas en encuestas sobre abortos. Véase también Rockwood, Sangster y Dillman, 1997; Holleman, 1999; Chang y Krosnick, 2003; Schwarz, Knäuper y Oyserman, 2008.

sido confeccionando tomando como referencia distintos cuestionarios usados habitualmente en el campo de la investigación sociopolítica y electoral.

b) Hipótesis

La hipótesis de partida del presente estudio es que la redacción y el orden tanto de las preguntas como de las categorías de respuesta pueden ejercer una influencia o introducir un sesgo en la información generada. A continuación se enumeran las variables en las que se han introducido tales cambios. En el apartado 4 se presentan los resultados finales.

- Hipótesis 1: Valoración de la situación general actual de España y de su evolución en el último año.

Se ha procedido a cambiar el orden de las categorías de respuesta en las variables de situación general de España y evolución del país. En concreto, las categorías de respuesta han sido invertidas, de tal manera que en el GC aparecen ordenadas de la más positiva a la más negativa, y en el GE a la inversa.

La hipótesis planteada es que la aparición en primer lugar de las categorías negativas (y no de las positivas, como es usual) podría orientar las respuestas hacia una peor valoración en ambas cuestiones. Este supuesto se basa en la idea de que el *efecto primacía*, esto es, que una opción será más elegida si se presenta en primer lugar, tendrá relevancia en una pregunta de estas características.

- Hipótesis 2: Valoración de la política, la familia y los amigos.

En este caso se ha cambiado el de orden de presentación de estos tres ítems que los sujetos debían valorar de 0 a 10 según la importancia que tuviesen para su vida.

Más específicamente, la hipótesis principal se basa en el cambio de la valoración de la variable “*política*” según esta sea preguntada al inicio del bloque, o bien tras las variables “*familia*” y “*amigos*”. En concreto, es de esperar que esta variable disminuya su valoración al ser enunciada tras “los amigos” y “la familia”, al tiempo que éstas pueden ver incrementadas sus puntuaciones al ir tras “la política”. La comparación que se establece se basa en que, en estudios previos, la política acostumbra a salir mucho peor valorada que las otras variables de carácter más personal³.

- Hipótesis 3: Valoración de líderes políticos: Zapatero, Rajoy e Ibarretxe.

Otra de las alteraciones realizadas tiene que ver con la hipótesis de que la valoración de un líder será más favorable si la persona por la que se pregunta antes es menos valorada, ya que por comparación se le asignará una mayor puntuación.

En el presente estudio, la comparativa se pretende realizar sobre el cambio en la valoración de un líder político habitualmente poco valorado, Juan José Ibarretxe, dependiendo de si se pregunta previamente por Zapatero y por Rajoy o el lehendakari vasco aparece en primer lugar. Según lo expuesto, y dadas las bajas

³ En el barómetro del CIS de enero de 2008, se valoraban estos aspectos en la misma escala de 0 a 10, con los siguientes resultados: la familia: 9,57; los amigos: 8,09; la política: 4,10.

puntuaciones de este último en las encuestas⁴, es de esperar que la valoración de Ibarretxe aumente al aparecer en primer lugar. Del mismo modo, se espera que la nota media de Rajoy sea superior si se pregunta por él después de Ibarretxe que si va detrás de Zapatero.

- Hipótesis 4: Índice de popularidad del presidente del gobierno.

El índice de popularidad, una variable que se suele presentar en todos los estudios de forma similar, también ha sido objeto de cambio. En este caso, la variación reside en la modificación de la redacción más habitual (“¿Aprueba usted la gestión actual de José Luis Rodríguez Zapatero como Presidente del Gobierno de España?”) por otra de carácter experimental (“¿Considera que la gestión actual de José Luis Rodríguez Zapatero como Presidente del Gobierno de España es más bien positiva o más bien negativa?”)

La hipótesis es que el cambio en la redacción de la pregunta puede conllevar una modificación del índice de popularidad, a pesar de que en ambos casos se esté aludiendo a un concepto muy similar. De hecho, cabría plantearse, en el caso de que se hallen diferencias, qué redacción recoge mejor el espíritu del concepto que se pretende medir. La más habitual tiene una clara connotación “escolar”: se está aprobando o desaprobando (suspendiendo) al presidente, por lo que la valoración tendría una clara traslación numérica. La redacción alternativa parecería tender más, *a priori*, a la realización de una “balance de la gestión” y la persona encuestada podría enfrentarse a una disyuntiva del siguiente tipo: *en conjunto, sopesando las cosas buenas y las malas, puedo concluir que la gestión es más bien positiva o más bien negativa.*

- Hipótesis 5: Intención de voto y simpatía política.

Una de las hipótesis finales es relativa al cambio de orden de dos preguntas que se emplean habitualmente en los cuestionarios de carácter sociopolítico: de un lado, la intención de voto, y de otro la simpatía política. La primera de estas variables es clave en el análisis dada su relevancia en los estudios electorales.

- Hipótesis 6: Autoubicación ideológica

Finalmente, el último cambio supone la alteración de las categorías de respuesta de la variable de autoubicación ideológica. La hipótesis de partida es que la asignación del valor “1” como punto más a la derecha, en lugar de a la izquierda, como es usual, *centraría* los resultados obtenidos. Esta afirmación se basa en un efecto psicológico, según el cual situarse en el punto “10” supone un valor más extremo que hacerlo en el “1”. Esto, unido a que la población que se siente más próxima a la derecha política suele ser más reacia a expresar su ideología, puede suponer una mayor participación de este grupo, y, por lo tanto, un mayor centramiento de la autoubicación media.

⁴ Según el pulsómetro de la SER del 26 de junio de 2008, la valoración de Zapatero es de 4,81, la de Rajoy 4,60, y la de Ibarretxe 2,51.

Metodología

Para la comprobación de las hipótesis se ha procedido a la aplicación de la técnica *split ballot*, que puede ser definida como un procedimiento por el cual una muestra es dividida en dos mitades, administrándose un cuestionario ligeramente diferente a cada una de ellas.⁵

El tamaño de la muestra total es de 602 unidades, dividido en dos mitades de 301 personas cada una. La información ha sido obtenida mediante encuesta telefónica asistida por ordenador (sistema CATI), para lo cual se ha utilizado un listado telefónico aleatorizado del ayuntamiento de A Coruña.

El muestreo es de tipo aleatorio estratificado, con aplicación de cuotas de sexo y edad en cada una de las muestras. Las desviaciones entre la afijación teórica y la real de cada muestra fueron corregidas mediante un equilibrio muestral posterior, de tal forma que cada una de ellas fuera equiparable en términos de sexo y edad. A continuación, se presenta la ficha técnica del estudio.

Cuadro 1.
Ficha técnica

Ámbito:	Municipio de A Coruña
Universo:	Población residente en el ayuntamiento de A Coruña con 18 y más años y línea telefónica fija en su hogar
Nivel de confianza:	95,45%
Probabilidad:	$p = q = 50\%$ (hipótesis de máxima indeterminación)
Tamaño total de la muestra:	602 personas (301 en cada submuestra)
Error máximo admitido:	5,76% en cada submuestra (4,07% global)
Fecha de ejecución:	29 de julio a 20 de agosto de 2008
Tipo de entrevista:	Telefónica asistida por ordenador (sistema CATI)
Tipo de muestreo:	Muestreo aleatorio estratificado por cuotas combinadas de edad y sexo con equilibrio muestral posterior

Resultados

Los resultados del estudio se presentan en cuatro grandes bloques:

1. Valoración de la situación actual general de España y su evolución reciente.
2. Grado de importancia de la política, la familia y los amigos.
3. Valoración de líderes políticos.
4. Intención de voto, simpatía y autoubicación ideológica.

⁵ “A procedure where a sample is divided into two halves and each receives a slightly different questionnaire”. ESOMAR: Market Research Glossary. Disponible online: <http://www.esomar.org/index.php/glossary-s.html>.

Con el fin de verificar si las diferencias encontradas en las distintas variables se deben al azar o resultan significativas, se procederá a aplicar pruebas estadísticas de contraste de hipótesis. En el caso de variables nominales u ordinales se aplicará el estadístico V de Cramer (menos sensible al tamaño muestral que la prueba de chi cuadrado), mientras que en el caso de variables numéricas se realizará una prueba T de contraste de medias.

- b) Valoración de la situación actual general de España y su evolución reciente
- Valoración de la situación general de España

Esta variable estaba redactada del siguiente modo “*Para empezar, refiriéndonos a la situación actual general de España, ¿cómo la calificaría Vd...*”

GC: ... *muy buena, buena, regular, mala o muy mala....?*”.

GE: ... *muy mala, mala, regular, buena o muy buena?*”.

Tabla 1.
*Tabla de contingencia Estudio * Situación actual general España*

		Muy buena	Buena	Regular	Mala	Muy mala	Total
Control	Recuento	2	31	161	79	28	301
	% de fila	,7%	10,3%	53,5%	26,2%	9,3%	100,0%
Experimental	Recuento	2	27	130	92	50	301
	% de fila	,7%	9,0%	43,2%	30,6%	16,6%	100,0%
Total	Recuento	4	4	58	291	171	78
	% de fila	,7%	,7%	9,6%	48,3%	28,4%	13,0%

Como se puede apreciar, parece existir una tendencia a valorar peor la situación general actual del país si se presentan primero las opciones de respuesta negativas. Las diferencias son tales que alcanzan el nivel de significación estadística con la aplicación del estadístico V de Cramer: su valor es de ,134 con una significación de ,029.

Estas diferencias se aprecian mejor si se procede a recodificar las categorías de respuesta en tres opciones: en el GC el 35,5% responden “mal” o “muy mal”, porcentaje que se eleva hasta el 47,2% en el GE.

Este efecto es mucho más intenso que el contrario, ya que las valoraciones de “muy buena” y “buena” son similares en ambos grupos (GC: 11,0%, GE: 9,6%). El cambio más destacable en el GC es la mayor importancia de la valoración “regular” (53,5%), 10 puntos superior a la que alcanza en el GE (43,2%). Las diferencias son significativas estadísticamente: el valor de la V de Cramer es ,119 ($p = ,014$).

- Percepción de la evolución reciente

Las preguntas estaban redactadas del siguiente modo: “*Y, ¿cree Ud. que la situación actual general del país es...*”

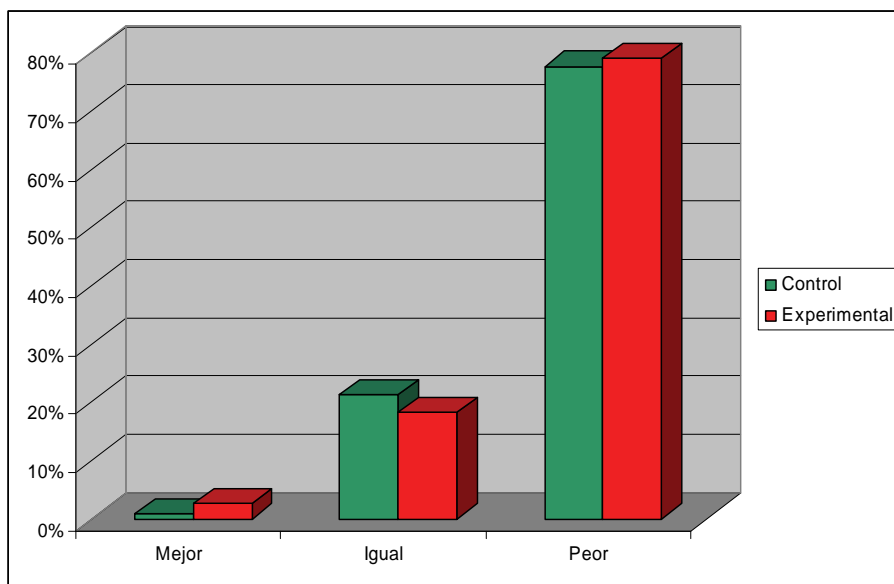
GC: ... *mejor, igual o peor que hace un año?*”

GE: ... *peor, igual o mejor que hace un año?*”

No ocurre lo mismo, sin embargo, cuando se pregunta por la percepción de la evolución de la situación de España en el último año (Ilustración 1). En ese caso, el cambio en el orden en que se presentan las categorías de respuesta apenas producen diferencias dignas de mención. El estadístico V de Cramer arroja un valor de ,071 con una significación de ,225.

Tal vez el hecho de que en la variable anterior (situación general de España) se incluyeran 5 categorías y en esta tan solo 3 haya influido en este resultado. Pudiera ser que el hecho de mencionar de forma consecutiva dos categorías de cariz negativo en el primer caso (“muy mala”, “mala”) hubiera tenido un impacto mayor en la percepción de los encuestados y encuestadas que mencionar tan solo una en el otro (“peor”).

Ilustración 1.
Evolución de la situación general de España



c) Grado de importancia de la política, la familia y los amigos

En este segundo bloque el objetivo era analizar si un cambio en el orden de presentación de tres ítems podía conllevar alteraciones en los resultados de cada una de ellas. La pregunta planteaba la importancia que de 0 a 10 tienen dichos aspectos en la vida de la persona encuestada. El orden de presentación era el que se detalla seguidamente:

Orden	GC	GE
1	Familia	Política
2	Amigos	Amigos
3	Política	Familia

A continuación se muestran los resultados.

Tabla 2.

Grado de importancia (0-10) de familia, amigos y política: estadísticos descriptivos

	Estudio	Media	Desviación típica	Significación
Política	Control	4,56	2,921	,011
	Experimental	5,16	2,842	
Amigos	Control	8,55	1,550	,243
	Experimental	8,70	1,553	
Familia	Control	9,52	1,245	,038
	Experimental	9,71	,987	

Como se puede observar, y tal y como se hipotetizó en el apartado , el cambio de orden afecta a las valoraciones. Más concretamente, la hipótesis principal afirmaba que la puntuación del ítem “política” sería más alta si se situaba en el primer lugar que si ocupaba el último lugar.

Se aventuraba para ello la interpretación de que, al carecer de la referencia de la puntuación dada a aspectos tan importantes en la vida de las personas como la familia o los amigos, los encuestados y encuestadas tenderían a dar una puntuación superior que si contaban con tales elementos de comparación. Los resultados parecen respaldar esta interpretación, pues en el caso de ser leída en primer lugar la política alcanza una valoración media de 5,16, lo que supone 6 décimas más que si es colocada en último lugar (4,56). Estas diferencias son significativas desde un punto de vista estadístico ($p = ,011$).

d) Valoración de líderes políticos

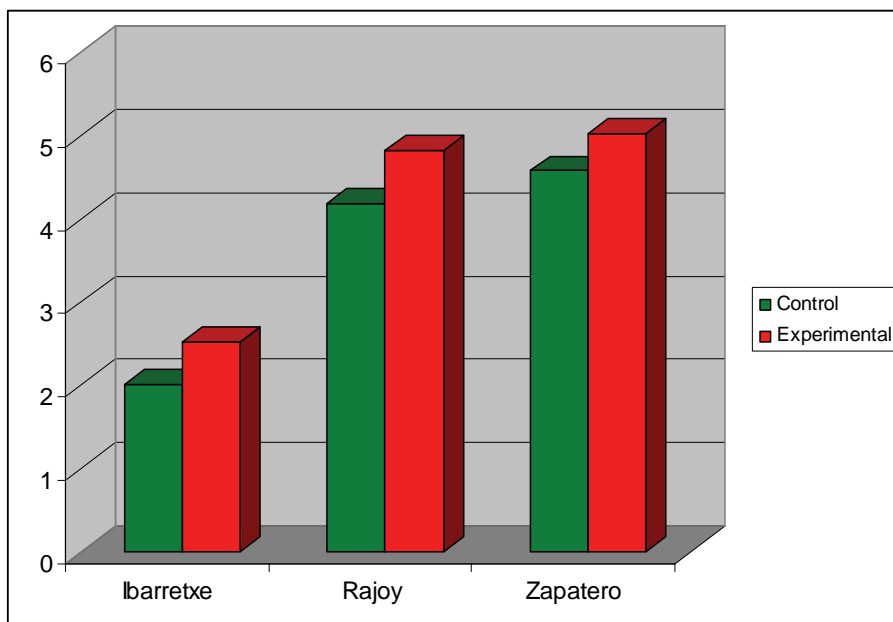
- Valoración de Zapatero, Rajoy e Ibarretxe

En este bloque se ha testado algo similar al anterior, pero en este caso se cambiaba el orden de presentación de distintos líderes políticos y se pedía que valoraran su actuación política de 0 a 10. Estas preguntas estaban filtradas de tal forma que sólo aquellas personas que conociesen al líder en cuestión procediesen a valorarlo. El orden de presentación fue el siguiente.

Orden	GC	GE
1	Zapatero	Ibarretxe
2	Rajoy	Rajoy
3	Ibarretxe	Zapatero

Desde un punto de vista teórico, se esperaba que la valoración de Ibarretxe fuera superior en el segundo caso, cuando los encuestados y encuestadas aún no tenían como referencia la nota dada a Rajoy y Zapatero.

Ilustración 2.
Valoración media líderes políticos



Como se puede observar, la nota dada a Ibarretxe es 5 décimas superior cuando es presentado en primer lugar (2,53 frente a 2,02).

Sin embargo, el efecto en el cambio del orden no se limita en este caso a una única categoría de respuesta, ya que las valoraciones de Rajoy y Zapatero también se ven afectadas. Así, ambos aumentan sus puntuaciones cuando se pregunta por su actuación política después de Ibarretxe.

En conclusión, los tres candidatos obtienen una mayor nota media si se emplea el orden del cuestionario experimental. En los tres casos, además, estas diferencias son significativas ($p < .05$).

- Índice de popularidad de Rodríguez Zapatero

En cuanto al índice de popularidad de Zapatero, no se ve excesivamente afectado por la diferente redacción de las preguntas:

GC: “¿Aprueba usted la gestión actual de José Luis Rodríguez Zapatero como Presidente del Gobierno de España?”

GE: “¿Considera que la gestión actual de José Luis Rodríguez Zapatero como Presidente del Gobierno de España es más bien positiva o más bien negativa?”

Tabla 3.
Tabla de contingencia Estudio * Índice de popularidad de Zapatero

		Aprueba / Más bien positiva	Desaprueba / Más bien negativa	Total
Control	Recuento	141	154	295
	% de Estudio	47,8%	52,2%	100,0%
Experimental	Recuento	152	132	284
	% de Estudio	53,5%	46,5%	100,0%
Total	Recuento	293	286	579
	% de Estudio	50,6%	49,4%	100,0%

Como se puede apreciar, y a pesar de medir un concepto similar a través de dos variables redactadas de un modo distinto, las diferencias son pequeñas y, en todo caso, no resultan significativas desde un punto de vista estadístico. El valor de la V de Cramer es ,057 con una significación de ,168.

- e) Intención de voto, simpatía y autoubicación ideológica
- Intención de voto

Una hipótesis similar se testaba en el presente bloque. En este caso, se trataba de analizar la influencia que podría tener el intercambio de lugar entre las variables “intención de voto” y “simpatía política”. El orden “ortodoxo” sitúa en primer lugar la intención y seguidamente la simpatía, que es el que se ha seguido en el GC. En el GE se ha invertido esta colocación. Los resultados aparecen recogidos en el Tabla 4. El estadístico V de Cramer arroja un valor de ,158 con una significación de ,013.

Como se puede apreciar, el orden experimental parece producir un efecto de aumento de respuesta a partidos y un descenso del porcentaje de indecisión y abstención. Sin embargo, en el Partido Popular se produce un efecto inverso y disminuye el porcentaje de personas que declaran que lo votarían.

Podría interpretarse como causa de este efecto que la simpatía por un partido constituye un rasgo identitario más perdurable que la efímera conducta del voto. Como tal, cabría pensar que los y las votantes de partidos más a la izquierda no sólo tienen una menor dificultad en reconocer abiertamente este hecho, sino que hacen de este elemento un rasgo en cierto modo definitorio de su propia identidad.

Entre quienes se declaran votantes del Partido Popular podría ocurrir lo contrario. Les resultaría más fácil declarar su intención de voto concreta (“si mañana

se celebrasen elecciones”) que reconocer una más estable disposición de tipo identitario, tal y como implica la simpatía por un partido.

Tabla 4:
*Tabla de contingencia Estudio * Intención de voto (recodificada)*

		PSOE	PP	BNG	Otros	No votará	No sabe	Total
Control	Recuento	77	63	15	42	41	53	291
	% de fila	26,5%	21,6%	5,2%	14,4%	14,1%	18,2%	100,0%
Experimental	Recuento	97	52	26	53	32	31	291
	% de fila	33,3%	17,9%	8,9%	18,2%	11,0%	10,7%	100,0%
Total	Recuento	174	115	41	95	73	84	582
	% de fila	29,9%	19,8%	7,0%	16,3%	12,5%	14,4%	100,0%

En definitiva, esta interpretación descansa en la idea de que una vez se declara la simpatía parece más sencillo reconocer el sentido del voto. Y al contrario, para quien resulta más difícil admitir por qué partido siente simpatía o carece de una simpatía definida con claridad, se complica todavía más reconocer a continuación su intención de voto.

- Autoubicación ideológica

La variable autoubicación ideológica fue sometida a un cambio diferente. En este caso, se optó por invertir el orden tradicional de la escala de 1 a 10.

GC: 1 punto más a la izquierda y 10 punto más a la derecha

GE: 1 punto más a la derecha y 10 punto más a la izquierda

El resultado esperado de este cambio de orden, tal y como se explicó en el apartado , era un menor escoramiento de la media hacia la izquierda y, por tanto, una mayor centralidad. Como es bien conocido, la escala estándar (la que usa habitualmente el CIS, por ejemplo) carece de punto intermedio, ya que el 5 (punto aparentemente central) está situado hacia la izquierda. A pesar de ello, la percepción habitual entre la población es que el 5 equivale al centro político.

Como consecuencia de ello, tal y como afirma Pilar del Castillo, “este indicador está sesgado a la izquierda en una medida difícil de evaluar”. No obstante, se justifica la utilización de la escala de 1 a 10 para evitar “probablemente (...) las connotaciones negativas de asignar la puntuación 0 a la izquierda” (Castillo, 2004:177).

Sin embargo, similares connotaciones negativas podría tener la elección del 10 en el caso de la derecha, ya que supone para la persona entrevistada situarse en el único número de dos dígitos de la escala y, por tanto, la sensación de estar en un lugar extremo.

En todo caso, el efecto que a priori parece más importante es el que se deriva de la utilización de una escala par, carente por tanto de punto intermedio, y en la que el punto central percibido se sitúa escorado hacia la izquierda.

En este sentido, el cambio de orden de la escala parece una estrategia interesante para poner a prueba las reflexiones precedentes. Los resultados se presentan en los siguientes cuadros e ilustraciones.

Tabla 5.
Autoubicación ideológica: estadísticos de grupo

Estudio	N	Media	Desviación típ.
Control	286	4,67	1,904
Experimental	281	5,97	1,965

Tabla 6.
Autoubicación ideológica: distribución en cada punto de la escala (% horizontales)

	Autoubicación política										Total
	1	2	3	4	5	6	7	8	9	10	1
Estudio (%)	7,0	5,3	11,6	14,7	39,6	9,1	6,3	2,5	,7	3,2	100,0
Control (%)	3,6	1,1	3,2	6,0	32,4	15,3	18,9	10,7	1,8	7,1	100,0
Experimental (%)	5,3	3,2	7,4	10,4	36,0	12,2	12,5	6,5	1,2	5,1	100,0

Como se puede apreciar, el cambio en el orden de la escala provoca una menor concentración en el punto central percibido (5) y una elevada frecuencia de respuestas en los puntos 7 y 6. No obstante, y a pesar de todo, el 5 continúa siendo la moda con total claridad.

En cierto modo, estos resultados ponen en cuestión la idea de que la escala de autoubicación tradicional está sesgada hacia la izquierda porque el punto central percibido cae hacia la izquierda.

Si aceptamos la interpretación de que la gente percibe el 5 como el centro político, y que quienes mencionan el 5 consideran que se están situando exactamente en el centro político, esto también tendría que ser válido en el caso de cambiar el orden de los polos. Y, por tanto, la media debería en este segundo caso situarse también en torno a dicha cifra.

Sin embargo, en el orden experimental la media se acerca claramente al 6: se aleja en casi 1 punto del 5, mientras que en el orden habitual sólo se aleja en 0,33.

Puede existir un efecto psicológico-espacial detrás de este hecho. Aceptando el hecho de que la ideología de la población no sigue una distribución estadística normal ligeramente sesgada hacia la izquierda, el cambio en el orden de los polos estaría dejando un mayor espacio de autosituación a la derecha de la escala para una mayor proporción de población (la que se considera de izquierdas).

Esto es así porque en el orden normal sólo hay 4 puntos hacia la izquierda (asumiendo que el 5 es el centro a efectos de percepción), mientras que con el cambio quedan 5. De esta forma, para alguien de centro izquierda puede resultar

incómodo situarse en el 1, el 2 ó incluso el 3 (demasiado pegado a los extremos), de ahí la tendencia de situarse en el 4.

Por el contrario, cuando la izquierda política pasa a ocupar el espacio del 6 al 10 existe mayor margen para los individuos. Por ejemplo, se puede elegir el 7 y estar todavía muy alejado del extremo (hay tres puntos más hacia la derecha de la escala). De hecho, el 7 es el punto más mencionado en la escala experimental, sólo por detrás del 5.

Sea cual fuere la interpretación o las objeciones que se puedan poner, queda un interrogante en el aire. Si aceptamos el hecho de que el 5 es el punto central percibido, y que este valor sea la moda en ambos grupos parece respaldar esta idea, quedaría por responder la siguiente pregunta: ¿por qué en el GE se sitúa significativamente menos gente en el 5 que en el GC?

En cualquier caso, estas diferencias se difuminan por completo si en lugar de calcular la media procedemos a transformar la variable numérica en categórica, tal y como se hace habitualmente (ver Tabla 7). En este caso, el estadístico V de Cramer arroja un valor de ,063 con una significación de ,689.

Tabla 7.

*Tabla de contingencia Estudio * Autoubicación ideológica rec. (5 categorías)*

		Izquierda	Centro izquierda	Centro	Centro derecha	Derecha	Total
Control	Recuento	35	75	139	25	12	286
	% de fila	12,2%	26,2%	48,6%	8,7%	4,2%	100,0%
Experimental	Recuento	25	83	134	27	13	282
	% de fila	8,9%	29,4%	47,5%	9,6%	4,6%	100,0%
Total	Recuento	60	158	273	52	25	568
	% de fila	10,6%	27,8%	48,1%	9,2%	4,4%	100,0%

Conclusiones

La principal conclusión del presente estudio es que el orden sí importa y puede alterar de forma significativa los resultados de alguna variable. De forma más concreta, las principales conclusiones que se extraen de este estudio pueden quedar sintetizadas tal y como se recoge a continuación:

La variable valoración de la situación general actual de España se ve afectada por el orden en que se presentan las categorías de respuesta. Es más negativa si las opciones de respuesta negativas se leen en primer lugar y a la inversa.

Esto plantea el interrogante de cuál de las dos formas de presentar las categorías de respuesta es más adecuada y la respuesta pueda ser que, probablemente, ambas lo sean. La forma de neutralizar el sesgo que se produce no pasa por optar por una u otra, sino por *evitar que en todos los cuestionarios se lea el mismo orden*. Esto se puede conseguir aleatorizando o rotando el orden en que presentan las categorías de respuesta, lo cual constituye ya práctica habitual en las investigaciones que se realizan usando ordenador (métodos CATI, CAPI) o PDA (método HAPI). Sin embargo, en caso de encuestas presenciales administradas en papel, la efectiva rotación de bloques de preguntas y categorías de respuesta queda a merced de los encuestadores y encuestadoras, por lo que no hay garantías de que tales rotaciones o aleatorizaciones se produzcan efectivamente. Esto implica, por tanto, un potencial problema para comparar datos procedentes de encuestas aplicadas de forma distinta en variables sensibles a cambios de orden.

La variable evolución de la situación general del país a lo largo del último año no experimentó las mismas variaciones que en el caso anterior. La presentación de la opción de respuesta negativa en primer lugar no implicó cambios significativos en los resultados. El hecho de que sólo hubiera una opción negativa, y no dos como en la variable anterior, puede explicar que en este caso el *efecto de primacía* quedase amortiguado.

Los resultados han mostrado con cierta contundencia que el orden en que se presentan determinadas variables interrelacionadas puede afectar a los resultados individuales de cada una de ellas. Más concretamente, cuando existe un factor que tiene una peor consideración social *probada* (a partir de estudios previos) que otros, el hecho de presentarlo en primer lugar conlleva una mejor valoración que si es presentado en último lugar. Con ello parece confirmarse que este tipo de valoraciones son en cierto modo contextuales y que los individuos construyen sus propias respuestas a partir de las respuestas que ya han dado, empleado las preguntas anteriores como referencias. Los resultados del factor “política” cuando se valora antes y después de “los amigos” y “la familia” y la valoración dada a Ibarretxe antes o después de Zapatero y Rajoy parecen sustentar esta interpretación.

Los resultados muestran que el cambio en el orden puede tener una mayor relevancia que el cambio integral de redacción de una variable, como ejemplifica el caso del índice de popularidad. Como se ha visto, dos operativizaciones totalmente diferentes de un mismo concepto ofrecen resultados comparables.

Uno de los resultados más sorprendentes tiene que ver con los efectos producidos por el cambio de orden de las variables intención de voto y simpatía

política. La ubicación de esta última variable en primer lugar trajo consigo una mayor tasa de respuesta global en la variable intención de voto. Sin embargo, con ese cambio se disminuyó el porcentaje de gente que se declaró votante del PP, justamente el partido que habitualmente ya obtiene una infrarrepresentación en las encuestas electorales. Los resultados de este cambio, pues, abren interrogantes acerca de cuál puede ser la mejor estrategia de presentación de preguntas a la hora de estimar los resultados en unas elecciones.

Los resultados del cambio de orden en las opciones de respuesta de la variable autoubicación ideológica resultan sugerentes. Si se acepta el postulado según el cual la población encuestada percibe el 5 como centro ideológico en la escala del 1 al 10, los resultados sugieren que la media se escora más hacia la izquierda de lo que se había supuesto hasta el momento, en contra del mayor centramiento ideológico que se había hipotetizado.

Sin embargo, si se convierte la escala numérica en categórica, recodificando las categorías de respuesta en cinco opciones (izquierda, centro izquierda, centro, centro derecha y derecha) las diferencias se difuman casi totalmente.

La conclusión parece clara: para evitar los efectos derivados del sesgo de utilizar una escala que carece de punto central se puede proceder a categorizar siempre las respuestas. El inconveniente es perder “riqueza” en la información con la conversión de una variable numérica en categórica. Otra alternativa pasaría por construir una escala impar que eliminara los posibles sesgos de incluir el 0 y el 10. En este sentido, la alternativa del 1 a 9 merece ser considerada. En tal caso, coincidiría el centro percibido con el punto central geométrico.

En definitiva, los resultados del presente estudio muestran la influencia que el diseño y la redacción del cuestionario tienen en los resultados finales. Sugieren que es preciso dedicar más tiempo y reflexión a estructurarlo correctamente y a considerar de forma muy cuidadosa en qué orden se colocan las variables, qué categorías de respuesta se eligen y cómo se presentan. La existencia de refinados instrumentos técnicos, los cuales permiten la realización de diseños muestrales complejos y el control de diversas fuentes de error o inexactitud, no debe relegar a un segundo plano la labor de construcción del cuestionario.

En definitiva, merece la pena rescatar una conclusión a la que llegara Gallup hace ya más de 60 años y que ha ido siendo aparcada tal vez por su *incomodidad* para la investigación social presente: los resultados de las encuestas están más influidos por el diseño de las preguntas que por el diseño muestral⁶.

A modo de balance: limitaciones y futuras líneas de investigación

La influencia real que el cambio de orden de las variables simpatía e intención de voto puede tener es algo que merece la pena investigar en mayor profundidad. En el presente estudio se ha detectado un aumento significativo de la tasa de respuesta global cuando se pregunta primero por la simpatía, salvo en el caso del PP. Sería,

⁶ “Gallup concluded that survey results were influenced more by question design than by sample design” (citado en Glendall y Hoek, 2002).

por supuesto, necesario replicar este experimento antes de extraer conclusiones definitivas. No obstante, si se confirmara, ello podría abrir la puerta a la construcción de cuestionarios con itinerarios diferentes en función de la posible intención de voto de los encuestados y encuestadas. Para ello sería clave ser capaces de identificar variables previas que demostraran correlación con la intención de voto, de tal modo que en función de las respuestas previas se presentara primero la intención de voto o la simpatía. Cuáles pudieran ser estas variables y la mejora que dicho cambio pudiera tener en la estimación final de resultados es algo que queda abierto.

Otro aspecto que merece cierta atención futura es el de la autoubicación ideológica. La actual escala parece ser problemática, toda vez que no resulta “reversible”, lo cual tal vez apunte hace posibles problemas de validez. ¿Sería conveniente tender hacia una escala impar, que tuviese un punto central claro, para evitar estos inconvenientes?

En cuanto a las limitaciones del estudio, la más importante se deriva probablemente del tamaño muestral y el consiguiente margen de error (5,76% en cada mitad de la muestra), lo cual obliga a ser cautos. La extracción de conclusiones más robustas exigiría la realización de una encuesta con un margen de error menor y, probablemente, la aplicación de pruebas estadísticas más complejas que podrían completar las empleadas aquí.

Referencias

- Belson, W.A. 1986 “Validity in survey research” *Gower Publishing Co.*, Londres
- Bishop, G; Smith, A. (2001) “Response order effects and the early Gallup Split-Ballots” *Public Opinion Quarterly* 65:479-505
- Carlson, C. S. 2005. “The Impact of question order on abortion surveys and who gets confused”. Comunicación presentada en la *Annual Meeting of The Midwest Political Science Association*, Chicago, Illinois.
- Castillo, P. 2004. “Nota metodológica sobre los indicadores del barómetro del CIS”. *REIS* 108/04:151-78.
- Chang, L.C.; Krosnick, J. 2003. “Measuring the frequency of regular behaviors: comparing the “typical week” to the “past week””. Comunicación presentada en la *Annual meeting of the American Association for Public Opinion Research*, Nashville.
- Converse, J.M, Presser, S. 1986 “Survey questions: handcrafting the standardised questionnaire” *Sage Publications*, Nueva Delhi, India.
- Culpepper, I.J.; Smith, W.R y Krosnick, J.A. 1992. “The impact of question order on satisficing in surveys”. Comunicación presentada en la *Midwestern Psychological Association Annual Meeting*, Chicago, Illinois.
- Díaz de Rada, V. (2002). "Influencia del orden de presentación de las respuestas en las elecciones de los entrevistados (por encuesta)", en *Política y Sociedad*, 39 (3): 691- 706.
- ESOMAR: “Market Research Glossary”. Disponible online: <http://www.esomar.org/index.php/glossary-a.html>.

- Dillman, D.A. et al 1995 "Effects of category order on answers in mail and telephone surveys" *Rural sociology* 60:674-87.
- Frankovic, K. 2008. "Why question order changes poll results: change the order of the questions in a poll, you change the outcome". *CBS News online* (disponible en <http://www.cbsnews.com/stories/2008/05/02/opinion/pollpositions/main4065521.shtml>).
- Gallup, G. 1947 "The quintadimensional plan of question design" *Public Opinion Quarterly*, 3: 385-393
- Glendall, Philip y Hoek, Janet. 2002. "A Question of Wording". *Marketing Bulletin* 1:25-36. Disponible en <http://www.sysurvey.com/tips/wording.htm>.
- Hippler, H.J.; Schwarz, N. 1987 "Response effects in surveys". en "Social Information processing and survey methodology", editado por Hippler, H.J., Schwarz, N, Sudman, S. *Springer Verlag* Pp. 102-21 New York
- Holbrook, Allyson L.; Krosnick, Jon A.; Moore, David y Tourangeau, Roger. 2007. "Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes". *Public Opinion Quarterly* 71 (3):325-348.
- Holleman, Bregje. 1999. "The nature of the forbid/allow asymmetry. Two correlational Studies". *Sociological Methods & Research* 28:209-244.
- Holleman, B.C. 2006 "The Meaning of "Yes" and "No". An Explanation for the forbid/allow asymmetry. Two correlational studies". *Quality&Quantity* 40:1-38
- Knäuper, B. (1999) "Age and Response order effects" *Public Opinion Quarterly* 63:347-370
- Krosnick, Jon A. 1989. "A review: question wording and reports of survey results: the case of Louis Harris and associates and aetna life and casualty". *Public Opinion Quarterly* 53 (1):107-13.
- Krosnick, J.A. & Schuman 1988 "Attitude intensity, importance, and certainty and susceptibility to response effects". *Journal of personality and social Psychology* 54 (6): 940-952
- 1999. "Survey Research". *Annual Review of Psychology* 50:537-67.
- Molenaar, N. 1982 "Response effects of "formal" characteristics of questions". In: W.Dijkstra and J. Van der Zouwen (eds), "Response Behaviour in the survey interview": *Academic Press*, pp 49-89 Londres
- Mc Allister, Ian y Wattenberg, Martin P. 1995. "Measuring Levels of Party Identification. Does Question Order Matter?" *The Public Opinion Quarterly* 59 (2):259-68.
- McFarland, S. G. 1981. "Effects of Question Order on Surveys Responses". *The Public Opinion Quarterly* 45 (2): 208-215.
- Mckee J. McClendon 1991 "Acquiescence and Recency Response-Order Effects in Interview Surveys" *Sociological Methods & Research*, 20 (1):60-103
- Quinn, SB, Belson, W.A., 1969 "The effects of reversing the order or presentation of verbal rating scales in survey interviews" *Survey Research Cent.* Londres.

- Rivers, D.C.; Meade, Adam W.; Fuller, W.L. 2008 "Examining Question and Context Effects in Organization Survey Data Using Item Response Theory" *Organizational Research Methods*
- Rockwood, Todd H.; Sangster, Roberta L. y Dillman, Don A. 1997. "The effect of response categories on questionnaire answers. Context and mode effects". *Sociological Methods & Research* 26:118-40.
- Schuman, H.& Presser, S. 1981/1996 "Questions and answers in attitude survey. Experiments on question forms, wording and context. *Thousand Oaks, CA: Sage Publications* (Originalmente publicado en Nueva York: Academic Press)
- Schwarz, Norbert; Knäuper, Bärbel y Oyserman, Daphna. 2008. "The psychology of asking questions". Pp 18-34 en *International Handbook of Survey Methodology*, editado por Leeuw, Edith D. de; Hox, Joop J. y Dillman, Don A. *Taylor & Francis Group*. Nueva York y Londres
- Schwarz, N.; Bless, H. 1992 "Constructing reality and its alternatives: assimilation and contrasts effects in social judgement" en "The construction of social judgement", editado por L.L. Martin; Tesser, A. *Hillsdale: Lawrence Erlbaum*, pp 217-45
- Schwarz, N., Hippler, H.J. (1987) "What response scales may tell your respondents: informative functions of response alternatives" en *Social Information, processing and survey methodology*" editado por Hippler, H.J., Schwarz, N., Sudman, S.. *Springer-Verlag*. Pp: 163-77 Nueva York
- Schwarz, N., Hippler, H.J.; Deutsch, B; Strack, F. (1985) "Response scales: effects of category range on reported behaviour and comparative judgements" *Public Opinion Quarterly* 49:388-95
- Schwarz, N., Hippler, H.J.; Noelle-Neumann, E. (1992) "A cognitive model of response -order effects in survey measurement" en *Context Effects in Social and Psychological Research*, editado por Schwarz, N, Sudman, S.: *Springer-Verlag* Pp 187-201 Nueva York
- Tourangeau, R; Rips, R. L.; Rasinki, K. 2000 "The psychology of survey response", *Cambridge University Press*, Cambridge.
- Tourangeau, R.; Singer, E.; Presser, S. 2003 "Context effects in attitude surveys. Effects in remote items and impact on predictive validity" *Sociological Methods & Research* 31:486-513