

Efectos de respuesta en cuestionarios con preguntas de actitud*

Vidal Díaz de Rada

Universidad Pública de Navarra. Departamento de Sociología
I-COMMUNITAS. Institute for Advanced Social Research
vidal@unavarra.es
ORCID: 0000-0002-9638-3741



Recepción: 20-07-2018
Aceptación: 14-04-2019
Publicación: 01-07-2020

Resumen

Existe un gran acuerdo según el cual el orden de presentación de las categorías de respuesta influye en las respuestas de los entrevistados, aunque los resultados son diversos y no muy concluyentes. Esta investigación busca contribuir a esta área de trabajo utilizando una investigación telefónica con un cuestionario con respuestas administradas en orden diferente y preguntas *descompuestas* en dos partes.

Los resultados desvelan que la encuesta telefónica presenta efectos de primacía y de recencia. Los primeros afectan a las dos primeras categorías, y los segundos únicamente a la penúltima categoría de respuesta. Los efectos de primacía son mayores cuando las categorías se muestran en orden inverso y en un formato de administración en dos preguntas. Los efectos de recencia, por su parte, en la penúltima categoría, son más frecuentes en el orden favorable y en el formato convencional.

Palabras clave: primacía; recencia; efectos de respuesta; efecto *satisficing*

Abstract. *Response effects in questionnaires with attitude questions*

Although research on the topic has produced mixed and inconclusive findings, it is largely agreed that the presentation order of response categories influences the answers given by survey respondents. This study aims to contribute to this area of research by examining the effects of reversing the negative/positive order of the response options and branching the questions in a telephone survey.

The results show the telephone survey to be influenced by primacy effects in the first two response categories and recency effects only in the penultimate category. Primacy effects appear more frequently when the response options are presented in reverse order with a two-way branching format, whereas recency effects are more frequent in the penultimate response category when the alternatives are ordered from positive to negative using the conventional (no branching) format.

Keywords: primacy; recency; response effects; *satisficing*

* Una versión previa de este trabajo fue presentada en la *I Jornada Intercongresual de Metodología de Investigación en Sociología*, organizada por la Federación Española de Sociología y la Associació Catalana de Sociologia.

Sumario

1. Antecedentes	Referencias bibliográficas
2. Aplicación a un caso: encuesta telefónica en una comunidad autónoma	Anexo 1. Temática de las preguntas de la encuesta telefónica en una comunidad autónoma española
3. Resultados	
4. Conclusiones	Anexo 2. Características de las cuatro muestras empleadas en la encuesta telefónica (porcentajes verticales)

La investigación sobre efectos de respuesta tiene una larga trayectoria y aumenta su intensidad en el ámbito anglosajón en la penúltima década del siglo xx. En ese contexto, la encuesta presencial predominaba los modos de recogida de información (Dillman et al., 2014), uso que decae en la penúltima década del siglo xx (Rossi et al., 1983; Groves, 2012; Wright y Mardsen, 2010; Singer, 2016). El estudio sobre esta temática ha experimentado un importante cambio a finales del presente siglo y se ha centrado fundamentalmente en los problemas de medida de las encuestas telefónicas, predominantes en el ámbito anglosajón y en el norte de Europa desde la década de los años noventa (Häder et al., 2012; Leeuw y Hox, 2015). Aunque en España se produce una implantación más tardía (Díaz de Rada, 2011), se trata del modo más utilizado en la investigación de opinión y mercados (Díaz de Rada y Portilla, 2015), y llega a ser la técnica dominante en determinados ámbitos como la estimación electoral (Díaz de Rada y Ayerdi, 2007).

Constatada la importancia de este tipo de encuesta, este trabajo trata de aportar información sobre los problemas de medición en modalidades que transmiten la información utilizando exclusivamente recursos orales, es decir, encuestas telefónicas. Concretamente, este trabajo busca descubrir los efectos propios del modo telefónico, originados, en primer lugar, por la ausencia de comunicación no verbal (al no ver al encuestador) y, en segundo lugar, por la mayor rapidez de la entrevista (Groves y Kahn, 1979), lo que dificulta su correcta realización. No menos importante es la posibilidad que tiene el entrevistado de realizar otra tarea mientras responde el cuestionario. Son efectos constatados en otros contextos (entre otros, Groves y Kahn, 1979; Ellis y Krosnick, 1999; Holbrook et al., 2003), pero totalmente desconocidos en España por la falta de investigaciones específicas. Esta constituye la primera aportación del trabajo,

De todos los posibles problemas de medición, este trabajo centra su atención en la influencia del orden en el que son ofrecidas las categorías de respuesta, para considerar hasta qué punto su disposición en el cuestionario aumenta o disminuye su elección. Un segundo objetivo es conocer la influencia de descomponer preguntas multicatóricas en dos preguntas con menos categorías, ya recomendada por los primeros trabajos específicos sobre encuestas telefónicas (entre otros, Dillman, 1978; Frey, 1989; Nicolaas et al., 2000).

Aunque una revisión de la literatura proporciona numerosos experimentos de cada uno de estos efectos (entre otros, Krosnick y Alwin 1987; Tourangeau y Rasinski, 1988; Krosnick y Presser, 2010; Gilbert, 2015), se aprecia una ausencia de trabajos que los consideren a ambos conjuntamente. Esta constituye la segunda aportación del trabajo.

Consciente de que en los últimos años se ha producido un enorme desarrollo de las encuestas autoadministradas a través de Internet, debe tenerse en cuenta que estas se basan, fundamentalmente, en recursos visuales (entre otros, Christian y Smyth, 2008; Christian et al., 2007). Aun cuando en ocasiones las encuestas autoadministradas incorporan recursos orales (encuestas web con voz), e incluso visuales-orales-paralingüísticos cuando utilizan entrevistadores virtuales,¹ se trata de una forma de comunicación totalmente diferente a la que tiene lugar en la encuesta telefónica (Dillman y Smyth, 2007; Maynard y Shaeffer, 2006). Esta forma de comunicación diferente, la presencia de efectos específicos en el modo autoadministrado (Dillman et al., 2005; Christian et al., 2007) y la gran utilización de la encuesta telefónica generan que la atención de este trabajo se centre en la localización de efectos de respuesta en las encuestas telefónicas.

El artículo tiene cuatro epígrafes. Comienza con un repaso teórico de la investigación sobre los efectos de respuesta, para proceder luego a explicar una aplicación concreta realizada en una comunidad autónoma española. En ese epígrafe se explica la fuente de datos, la situación experimental, las hipótesis del trabajo y las técnicas de análisis de datos. En el tercer epígrafe, se presentan los resultados, y se termina con las conclusiones y el listado de las referencias utilizadas.

1. Antecedentes

Las primeras investigaciones que estudiaron que la colocación de las respuestas de preguntas cerradas influya en las respuestas de los entrevistados fueron realizadas a mediados del siglo pasado (entre otros, Cantril, 1944; Payne, 1951). Desde entonces, un gran número de investigaciones han revelado la existencia de cambios en las respuestas producidos por la variación del orden en que las categorías de respuesta son presentadas (entre otros, Krosnick, 1991; Krosnick y Alwin, 1987; Tourangeau y Rasinski, 1988). Según Krosnick y Alwin, los efectos de la primera y la última opción están producidos porque los entrevistados no prestan la misma atención a las distintas opciones de respuestas ofrecidas. A pesar de que la respuesta puede guardar relación con antiguas ideas ya asentadas en la mente, en el momento en que la pregunta es formulada

1. Hace referencia a la aparición de la cara de una persona en la pantalla del ordenador que lee las preguntas del cuestionario y recoge las respuestas. El encuestado tiene la sensación de que está siendo entrevistado por alguien, aunque es perfectamente consciente de que no es una persona, porque la imagen dista mucho de ser la de una persona real. Es, digamos, una especie de caricatura. Una explicación detallada puede consultarse en Zhang (2013) y Conrad (2019).

se realiza una rápida selección de la información disponible y la contestación tan solo refleja algunas de estas ideas —en línea con lo constatado por Zaller (2014) en sus estudios sobre la opinión pública norteamericana. De modo que el entrevistado es muy susceptible de ser influenciado por la información suministrada por el cuestionario.

Según esta propuesta, el surgimiento de los efectos de respuesta depende sobremanera del orden en que las categorías de respuesta son presentadas, y de si estas son expuestas en tarjetas o son leídas en voz alta a los entrevistados. Krosnick y Alwin (1987) reseñaron cómo las categorías de respuesta emplazadas al comienzo y al final del listado de opciones aumentan la oportunidad de ser escogidas y, por lo tanto, de provocar la aparición de efectos de la primera y la última opción. Desde su punto de vista, cuando las categorías de respuesta son escuchadas por los entrevistados, estos no tienen la oportunidad de seguir procesando la primera opción ofrecida. De esta manera, las categorías de respuesta ubicadas en los últimos lugares de la lista de opciones disponen de más tiempo para ser interiorizadas y, por lo tanto, tienen más probabilidades de ser escogidas, lo que provoca la aparición de efectos de la última opción (o recencia).

Pese a la solidez de esta argumentación y la comprobación de estos efectos en diversas investigaciones, numerosos autores creen que es necesario investigar más al respecto (entre otros, Dillman et al., 2014; Chang y Krosnick, 2010; Holbrook et al., 2007; Bishop y Smith, 2001). Este trabajo trata de responder a esta demanda comprobando hasta qué punto el orden de colocación de las categorías de respuesta genera cambios en su nivel de elección. Un segundo objetivo es verificar si la descomposición de preguntas en dos partes, recomendada para la administración telefónica (entre otros, Dillman, 1978; Frey, 1989; Krosnick y Berent, 1993; Nicolaas et al., 2000), produce cambios en las respuestas.

2. Aplicación a un caso: encuesta telefónica en una comunidad autónoma

La encuesta telefónica se basa exclusivamente en recursos orales, únicamente la voz y el tono de los interlocutores. Se trata de un proceso de entrevista caracterizado por la rapidez con la que se produce la dinámica pregunta-respuesta, y por un ritmo de entrevista totalmente controlado por el entrevistador (Lavrakas, 2010; Leeuw, 2008). Otro aspecto relevante para el tema tratado en este trabajo es la posibilidad de que el entrevistado esté realizando otras tareas cuando lleva a cabo la respuesta del cuestionario (cuadro 1). Estas situaciones

Cuadro 1. Características de las encuestas a través del teléfono.

Basada exclusivamente en recursos orales (voz)

Caracterizada por la rapidez con la que se produce la dinámica pregunta-respuesta

Ritmo de entrevista totalmente controlado por el entrevistador

Posibilidad de que el entrevistado esté realizando otras tareas mientras responde

Recursos orales y mayor rapidez = modalidad con gran presencia de efectos de recencia

Fuente: elaboración propia.

llevan a considerar como hipótesis que se trata de un modo con una gran prevalencia de efectos de la última opción o recencia.

2.a. Fuente de datos

La encuesta telefónica (a teléfonos fijos) empleada en la presente investigación tuvo como universo objeto de estudio a las personas de ambos sexos de 16 y más años residentes en hogares familiares. El universo de referencia fue el Padrón Municipal de Habitantes a 1 de enero del año 2015, formado por 539.745 personas. El tamaño muestral para estudiar de forma representativa este universo se estableció en 800 entrevistas, que, con un nivel de confianza del 95,5 % y $p = 0,5$, presentaba un error muestral máximo para datos globales de $\pm 3,5$ % (en el supuesto de un muestreo aleatorio simple).

Estratificado el universo según la zona de residencia y el hábitat de municipio, se utilizó un muestreo aleatorio estratificado. Las unidades primarias de muestreo (municipios) fueron seleccionadas de forma aleatoria con probabilidad proporcional a su tamaño, y las unidades secundarias (viviendas), con un muestreo sistemático de números de teléfonos. Las unidades últimas (individuos) se eligieron utilizando cuotas cruzadas de sexo y edad.

Se utilizó un cuestionario con 48 preguntas, de las que 10 eran sociodemográficas (que proporcionan otras tantas respuestas/variables) y 38 preguntas de contenido. El cuestionario fue pilotado utilizando entrevistas cognitivas y detección de los problemas de interacción en la entrevista (*behaviour coding*).² Además, se llevó a cabo un detallado proceso de formación de entrevistadores que incluyó una formación práctica de diez entrevistas antes de llevar a cabo la primera encuesta, siguiendo el proceso de formación y supervisión propuesto por Gwartney (2007) y Lavrakas (1993). El cuestionario se respondió en un promedio de 13 minutos.

La práctica totalidad de preguntas de contenido recogía opiniones y actitudes (nunca comportamientos) referidas a cinco temáticas: 1) valoración económica y política de la comunidad; 2) interés en la política y valoración de líderes; 3) conocimiento y valoración de cuatro instituciones relacionadas con el Parlamento, y 4) valoración de la actuación del gobierno autónomo (ver anexo 1). El análisis aquí planteado se refiere a estas cuatro temáticas, sin considerar la información sociodemográfica.

2. La dinámica «perfecta» de responder el cuestionario es que el entrevistador lea una pregunta, y el encuestado la escuche y reflexione sobre ella. Seguidamente, proporciona una respuesta que es anotada por el encuestador y se procede con la siguiente pregunta.

La técnica del *behaviour coding*, creada por Cannel (1992) y desarrollada posteriormente por Fower (1995 y 2011), considera todas las situaciones que se apartan de esta dinámica: preguntas al encuestador, solicitudes de repetición de la pregunta, no respuestas, etcétera. Se procede a contar cada una de esas situaciones y, cuando una pregunta obtiene un número elevado, es necesario reformularla.

Se ha utilizado porque se trata de una detección de errores objetiva, «libre» del efecto y parecer del entrevistador, así como por la eficacia contrastada por un gran número de investigadores (entre otros, Park y Lee, 2018; Zouden y Smith, 2004).

Aunque diversas investigaciones han localizado efectos de primacía y de recencia en preguntas dicotómicas (entre otros, Bishop y Smith, 1997; 2001; Krosnick et al., 2002; Holbrook et al., 2007), este trabajo se centra en las preguntas de más de tres categorías, lo que limita el análisis —de las 77 variables incluidas en el cuestionario— a 33, que cuentan con cuatro, cinco, diez y once categorías.

Pensando siempre en las limitaciones visuales de la encuesta telefónica, en la elaboración del cuestionario se puso máximo interés en plantear preguntas con enunciados breves y pocas categorías de respuesta, según las recomendaciones de Olson y Smyth (2015). Así, 13 variables tenían cuatro categorías, y otras 6 presentaban cinco categorías.³ Se trata de un tipo de preguntas que han sido definidas como «preguntas categorizadas de escala ordinal» (Azofra, 1999) o «preguntas multicotómicas» (García Ferrando y Llopis, 2015). Otros autores, como por ejemplo Visauta (1989), hablan de «abanico de respuestas» aludiendo a la forma en la que las alternativas de respuesta se presentan en el cuestionario. Considerando la disposición vertical de las categorías de respuesta, diferente a las escalas valorativas con opciones de respuestas en horizontal, en este trabajo se empleará el término «pregunta abanico». Además, en el cuestionario había también 14 variables con «escalas de valoración» (Azofra, 1999) entre 0 y 10.

Es importante precisar que ninguna pregunta ofrecía la categoría de respuesta «no sabe» ni «no desea responder», aunque a los entrevistadores se les señaló que diferenciaran cada situación.

Teniendo en cuenta el modo de transmisión de la información de la encuesta telefónica, se intentó, en todo momento, cumplir la recomendación de Fowler (1995) de no superar las 25-30 palabras por pregunta. Esta regla trató de seguirse en todas las preguntas, excepto en dos en las que no pudo aplicarse al disponer de 58 y 77 palabras.⁴ Cinco preguntas se sitúan cerca del límite superior.⁵ Es importante destacar que la práctica totalidad de las preguntas presentan un nivel de medida ordinal⁶ que reclama un menor esfuerzo cognitivo que la medida nominal (seleccionar una opción de un conjunto diverso de opciones).

2.b. Condiciones experimentales

El cuestionario empleado disponía de cuatro versiones, denominadas A1, A2, B1 y B2, que fueron aplicadas a cuatro muestras equivalentes del

3. Había preguntas que presentaban cinco categorías de respuesta, aunque solo cuatro eran leídas al entrevistado al no ofrecer la opción central.
4. Se trata de una pregunta con cuatro categorías sobre la capacidad del Gobierno autonómico para solucionar los problemas de la comunidad, y otra sobre la intención de votar o abstenerse en el caso de unas hipotéticas elecciones al día siguiente de cuando fue respondido el cuestionario.
5. Tres preguntas con cinco categorías de respuesta referidas a la frecuencia de consulta de noticias políticas en los medios de comunicación, que tienen 29 palabras. Dos escalas, 0-10, utilizadas para evaluar al Gobierno y a su presidente, emplean 28 palabras y 26, respectivamente. Del resto de preguntas, ninguna supera las 25 palabras.
6. Al interrogar al entrevistado sobre su nivel de acuerdo/desacuerdo, o su valoración «muy buena», «buena», «mala», «muy mala», de determinados aspectos.

mismo tamaño. Estas muestras presentan la misma distribución en cuanto a sexo, edad, nivel de estudios y relación con la actividad (anexo 2). La diferencia entre el cuestionario A1 y A2 está en el orden de lectura de las categorías de respuesta. En el cuestionario A1, las categorías comienzan en opciones favorables («muy bueno», «bien», etcétera) y terminan en opciones desfavorables («muy malo», «malo», etcétera). Se procede de esta forma con el fin de conocer la aplicación en nuestro contexto de las recomendaciones realizadas por algunos expertos de la conveniencia de colocar en primer lugar las respuestas positivas/favorables (entre otros, Christian et al., 2008; Stern et al., 2012; Tourangeau et al., 2013a; Lundmark et al., 2016). Estas categorías (A-B-C-D) son administradas de forma inversa en el cuestionario A2 (D-C-B-A). En el resto del trabajo, la primera administración será denominada como favorable y la segunda como desfavorable. El objetivo de esta forma de proceder es constatar hasta qué punto la administración oral genera un mayor recuerdo de las últimas categorías de respuesta, un mayor efecto de recencia, propuesta que constituye la primera hipótesis del trabajo.

La diferencia del tercer y cuarto cuestionario (versión B) con los anteriores es que cada pregunta se divide en dos: la primera pregunta es sobre tendencia o dirección, y la segunda sobre intensidad⁷ (Dillman, 1978; Frey, 1989). Se procede de esta forma siguiendo las recomendaciones de diversas investigaciones (entre otros, Krosnick y Berent 1993; Malhotra et al., 2009) que han comprobado que así aumenta la fiabilidad y la validez de las respuestas y se eliminan los efectos derivados del orden de las preguntas. Investigaciones más recientes han demostrado que esta división de cada pregunta en dos partes produce respuestas más extremas (Gilbert, 2015). Estos hallazgos han sido utilizados para la formulación de la segunda hipótesis.

Todo el cuestionario B se aplicó de esta forma, excepto las escalas valorativas entre 0 y 10, que fueron iguales a las utilizadas en la versión A. Este acercamiento paulatino a la pregunta será denominado como formato de administración en dos preguntas, mientras que la administración en una pregunta abanico —similar a los cuestionarios A— recibirá el nombre de administración convencional. La diferencia entre los dos cuestionarios B es que el B1 muestra primero las opciones favorables, mientras que el B2 comienza con las desfavorables.

Se planeaba como hipótesis que esta reducción en el número de categorías, al convertir una pregunta de cuatro opciones en dos preguntas (una de ellas dicotómica), eliminaría los efectos derivados del orden en el que se presentan las categorías de respuesta (tercera hipótesis). Ahora bien, la realización de un mayor número de preguntas —aunque más breves— reclama más tiempo para responder el cuestionario (Gilbert, 2015), lo que implica también un aumento del coste de la investigación —cuarta hipótesis.

7. Es decir, la primera busca conocer si el entrevistado está «de acuerdo» o «en desacuerdo» con una proposición, y la segunda plantea si está «bastante» o «muy» + respuesta anterior.

En definitiva, se cuenta con un diseño 2 x 2 en el que 400 cuestionarios presentan un orden de respuesta A-B-C-D, donde la mitad ha respondido preguntas de cuatro y cinco categorías y la otra mitad preguntas con dos y tres categorías que daban paso a otra pregunta con dos categorías. La otra mitad de la muestra responde las preguntas en la secuencia desfavorable (D-C-B-A), 200 en formato convencional y otros 200 en dos preguntas. Esta estructura se presenta en el cuadro 2, y a continuación se detalla la terminología empleada:

- Convencional favorable (columna izquierda, parte superior).
- Convencional desfavorable (columna izquierda, parte inferior)
- Dos preguntas favorables (columna de la derecha, parte superior).
- Dos preguntas desfavorables (columna de la derecha, parte inferior)

Cuadro 2. Diseño de la condición experimental.

Orden	Formato de administración de la pregunta		Total
	Formato convencional (Cuestionario B1)	Formato con dos preguntas (Cuestionario B2)	
Favorable: cuestionario A1, orden A-B-C-D	200	200	400
Desfavorable: cuestionario A2, orden D-C-B-A	200	200	400
Total	400	400	800

Fuente: elaboración propia.

La asignación de cada tipo de cuestionario a cada muestra fue realizada de forma aleatoria. Además, todos los entrevistadores aplicaron el mismo número de cuestionarios en cada muestra y entrevistaron a diferentes perfiles de personas.⁸ Se busca, de esta forma, reducir al máximo la posible influencia del encuestador.

Esto genera cuatro muestras equivalentes considerando sexo, edad, nivel de estudios y relación con la actividad (ver anexo 2). Con ambas situaciones experimentales, se busca comprobar las hipótesis que se proponen en el cuadro 3.

Terminado el trabajo de campo, las respuestas de los cuestionarios B fueron recategorizadas exactamente igual que las de los cuestionarios A para que pudieran ser adecuadamente comparados.

2.c. Técnicas de análisis de datos

Para localizar los efectos producidos por la colocación de las opciones de respuesta, se elaboró una variable para recoger el número de ocasiones en las que

8. Cada día se repasaban los rasgos de los encuestados por cada entrevistador, considerando edad, sexo y situación laboral, asignando rasgos diferentes para las entrevistas del día siguiente.

Cuadro 3. Hipótesis

- H1: La modalidad de recogida de información utilizada, al estar basada en recursos orales, tendrá un mayor número de efectos de recencia que de primacia
- H2: Las preguntas descritas en la hipótesis anterior proporcionan también respuestas más extremas
- H3: La reducción en el número de categorías que convierte una pregunta de cuatro o cinco opciones en dos preguntas eliminará los efectos derivados del orden en el que se presentan las categorías de respuesta
- H4: El aumento en el número de preguntas que supone la administración en dos preguntas reclama más tiempo para responder el cuestionario, lo que aumenta el coste del trabajo de campo
-

Fuente: elaboración propia.

cada entrevistado escogió la opción A o B, diferenciando si estas estaban situadas en primer o último lugar. Diversos expertos, entre otros Chang y Krosnick (2010), señalan que se produce recencia cuando las opciones A o B reciben más menciones en el grupo en el que se leen en último lugar, con respecto al grupo en el que se leen en primer lugar. Para la elaboración de la variable de primacia, se procede de forma similar, pero considerando la mayor elección de las opciones C y D cuando se leen en primer lugar.

Seguidamente, se considerará el número medio variables con efecto de primacia o de recencia respecto al total de variables consideradas. Para conocer si las diferencias respecto a la situación experimental⁹ es significativa, será utilizado el test de significación de la diferencia de medias, tal y como han procedido investigaciones similares realizadas en otros contextos (entre otros, Chang y Krosnick 2010; Holbrook et al., 2007; Holbrook et al., 2003; Krosnick et al., 2002; Bishop y Smith, 2001).

Con el fin de detectar el efecto de covariables sociodemográficas en las respuestas, se empleará la técnica de la regresión, considerando como término dependiente el número de preguntas con efectos (primacia y recencia) y, como independientes, la situación experimental (ver nota al pie nueve), el sexo, el nivel de estudios y la edad; estas últimas con notable influencia en la mayor parte de investigaciones sobre el tema (entre otros, Krosnick et al, 1996; Holbrook et al, 2003; Holbrook et al., 2007). Han sido introducidas también —como términos independientes— las interacciones entre el nivel de estudios-edad y las dos situaciones experimentales.

3. Resultados

La primera hipótesis plantea la existencia de un mayor número de efectos de recencia generados por la colocación de las categorías de respuesta y la forma en la que se transmite la información telefónica. En primer lugar, su comprobación necesita el cálculo del indicador. Como se ha señalado, se elabora considerando el número de veces que se responde utilizando las dos últimas categorías de la

9. Orden de administración de categorías y formato de administración de la pregunta.

Tabla 1. Número medio de preguntas con efectos de primacía y de recencia (y % respecto al total).

	Efecto de primacía	Efecto de recencia
Número medio	6,98	7,43
% respuestas respecto a las 33 prg. del cuestionario	21,1 %	22,5 %

Fuente: elaboración propia.

pregunta. Este varía entre 0 y 19, donde 0 implica que el entrevistado nunca ha elegido las dos últimas categorías de respuesta, y 19 que siempre las ha elegido. La media, 7,43 preguntas, considerando las 33 preguntas del cuestionario, supone que un 22,5 % de las preguntas del cuestionario han experimentado efecto de recencia (7,43 / 33). Ahora bien, a diferencia del trabajo de Chang y Krosnick (2010), que considera la elección de las dos últimas categorías como recencia, en el presente trabajo se diferencia entre una recencia extrema (elección de la última categoría) y recencia suave (elección de la penúltima categoría).

La hipótesis planteaba que la recencia será superior a la primacía, lo que comporta la elaboración del otro indicador seleccionando el número de veces que se ha respondido eligiendo las dos primeras categorías de la pregunta. Este proporciona un rango menor, al oscilar entre 0 y 17. El número medio de preguntas que han experimentado primacía es de 6,98, lo que implica que este efecto ha tenido lugar en un 21,1 % de las preguntas. La información presentada en la tabla 1 da cuenta de la existencia de efectos de primacía y de recencia.

La parte superior izquierda de la tabla 2 muestra que ofrecer las categorías de forma desfavorable (de menos favorable a más favorable) produce una mayor elección de la segunda categoría de respuesta, algo que sucede en un 14,7 % de las preguntas del cuestionario A2.¹⁰ Los efectos de recencia (parte superior

10. En un primer momento fueron analizadas las diferencias en el efecto de primacía «extrema» (primera categoría), que proporcionó un número medio de respuestas similares entre el orden favorable y el desfavorable. Esto mismo sucedió con el efecto de recencia en la última categoría. Por este motivo no han sido incluidos en esta tabla, en la que se han colocado únicamente las opciones con diferencia significativa.

Aquí se muestra el efecto completo del orden de las categorías:

	Efecto de primacía			
	Primera categoría		Segunda categoría	
	Favorable (ABCD)	Inverso (DCBA)	Favorable (ABCD)	Inverso (DCBA)
Nº medio de preguntas	2,59	2,35	4,14	4,90
% respuestas	7,8 %	7,1 %	12,5 %	14,7 %
	t 1,780; p 0,076		t -4,08; p 0,000	
	Efecto de recencia			
	Penúltima categoría		Última categoría	
	Favorable (ABCD)	Inverso (DCBA)	Favorable (ABCD)	Inverso (DCBA)
Nº medio de preguntas	4,85	3,87	3,12	3,03
% respuestas	14,7 %	11,7 %	9,5 %	9,2 %
	t 5,213; p 0,000		t 0,545; p 0,586	

Tabla 2. Influencia del orden de presentación de las categorías y del formato de administración.

	Orden de las categorías			
	Efecto de primacía (Segunda categoría)		Efecto de recencia (Penúltima categoría)	
	Favorable (ABCD-A1)	Desfavorable (DCBA-A2)	Favorable (ABCD-A1)	Desfavorable (DCBA-A2)
Nº medio de preguntas	4,14	4,90	4,85	3,87
% respuestas	12,5 %	14,7 %	14,7 %	11,7 %
	$t -4,08$; $p 0,000$		$t 5,213$; $p 0,000$	
	Formato de administración			
	Efecto de primacía			
	Primera categoría		Segunda categoría	
	Convencional-B1	Dos preguntas-B2	Convencional-B1.	Dos preguntas-B2
Nº medio de preguntas	2,05	2,90	4,18	4,84
% de respuestas	6,2 %	8,8 %	12,7 %	14,7 %
	$t 6,457$; $p 0,000$		$t -3,599$; $p 0,000$	
	Efecto de recencia			
	Última categoría		Penúltima categoría	
	Convencional-B1	Dos Preguntas-B2	Convencional-B1	Dos preguntas-B2
Nº medio de preguntas	2,89	3,26	5,52	3,17
% de respuestas	8,8 %	9,9 %	16,7 %	9,6 %
	$t 2,153$; $p 0,032$		$t -13,666$; $p 0,000$	

Fuente: elaboración propia.

derecha de la tabla 2) tienen lugar también en la categoría no extrema (penúltima), y aumentan en la administración favorable al afectar al 14,7 % de las preguntas del cuestionario A1, tres puntos por encima de la administración desfavorable (11,7). El aspecto más sorprendente de la primera parte de la tabla 2 es que ambos efectos se producen en la segunda y penúltima categoría. No en la categoría más extrema, sino en la que diferencia con menos rotundidad entre estar de acuerdo o desacuerdo.

Volviendo a la información de la tabla 2, obsérvese que la diferencia entre las categorías de cada tabla es mayor en el efecto de recencia, lo que implica que este efecto es superior a la primacía (valores $t 5,2$ y $4,1$ respectivamente).

Resulta difícil encontrar una lógica a estos resultados, que se caracterizan por una menor elección de respuestas extremas, mucho más cuando se considera que este es un efecto propio de las encuestas telefónicas (entre otros, Nicolaas et al., 2000; Ye et al., 2011). Ahora bien, esa mayor presencia de categorías intermedias coincide con los hallazgos localizados en investigaciones auto-administradas realizadas en otros contextos (entre otros, Schwarz y Hippler, 1987; Tourangeau et al., 2004; Smyth et al., 2007). Son investigaciones que muestran un efecto de aglomeración de las opiniones hacia el punto medio de la pregunta, al considerar que este representa la opción mayoritaria. Este des-

lizamiento hacia el punto medio, fácilmente reconocible en un formato visual, como la encuesta autoadministrada, supone un proceso de respuesta similar al efectuado cuando se utilizan estímulos visuales. Así, se han localizado hallazgos muy diferentes a lo destacado por la mayor parte de la literatura internacional.

El formato de administración presenta, de nuevo, efectos de primacía y de recencia. En la segunda parte de la tabla 2, se muestra por separado el número de preguntas donde han sido elegidas la primera y la segunda categoría, y ambas manifiestan la misma tendencia: hay una mayor presencia del efecto de primacía en las preguntas divididas en dos (cuestionario B2), con un promedio de 7,7 (2,9 + 4,84) preguntas, frente al 6,2 del formato convencional (2,05 + 4,18). La primacía afecta al 23 % (8,8 + 14,7) de las preguntas del cuestionario B2, en línea con la hipótesis de que este tipo de preguntas presenta respuestas más extremas (segunda hipótesis). Considerando que el formato en dos preguntas plantea —en este caso— cuestiones dicotómicas, la tasa obtenida es similar a la detectada en investigaciones realizadas en otros contextos (entre otros, Bishop y Smith, 2001; Holbrook et al., 2007).

En el efecto de recencia se produce una mayor elección de la última categoría en las preguntas en dos fases (cuestionario B2), mientras que en el formato convencional (cuestionario B1) hay una mayor elección de la penúltima categoría, con la presencia del efecto de aglomeración (o centralidad) anteriormente detectado.

Considerando conjuntamente las dos situaciones experimentales, el análisis presenta diferencias significativas en el efecto de primacía (agregadas ambas categorías) y en la recencia penúltima, y no así en la recencia extrema (última categoría). La parte inferior izquierda de la tabla 3 muestra la existencia de menor diferencia entre el orden de las respuestas que entre el formato de administración.¹¹ Esto indica que la influencia del orden de presentación de las respuestas es inferior al formato de administración, apareciendo situaciones en las que $\frac{1}{4}$ de las variables presentan efectos de respuesta (concretamente, primacía en el orden desfavorable con formato en dos preguntas). En la tabla derecha, se detecta la situación opuesta, ya que la aplicación desfavorable en dos preguntas reduce al mínimo el efecto de recencia penúltima: solo un 6,1 % presenta efectos de respuesta. La recencia penúltima aumenta notablemente en la aplicación convencional desfavorable.

Un análisis de toda la tabla revela que las categorías menos extremas son la segunda y la penúltima, las que presentan mayores diferencias en las dos situaciones experimentales, si bien en la primacía influyen ambas categorías.

Ante la sospecha de que esta situación fuera debida a diferencias muestrales, se lleva a cabo una regresión en que se considera como variable dependiente el efecto de primacía de las dos categorías, y como variables independientes las dos situaciones experimentales, el sexo, la edad, el nivel educativo, así como las interacciones edad-nivel educativo y orden de categorías-formato de administración.

11. Orden de respuesta $F = 5,53$, 1 g.l., signif. = 0,019. Formato de administración $F = 46,182$, 1 g.l., signif. 0,000.

Tabla 3. Influencia conjunta del orden de presentación de las categorías y del formato de administración (valores en porcentajes).

Orden	Formato					
	Primacia		Recencia última cat.		Recencia penúltima	
	Convencional-B1	Dos preguntas-B2	Convencional-B1	Dos preguntas B2	Convencional-B1	Dos preguntas-B2
Favorable-A1	19,9	20,9	9,1	9,8	16,3	13,1
Desfav.-A2	17,9	25,9	8,4	9,9	17,1	6,1
	f_{Orden} 5,53; p. 0,019		f_{Orden} 0,289; p. 0,591		f_{Orden} 37,544; p. 0,000	
	f_{Formato} 46,812; p. 0,00		f_{Formato} 4,618; p. 0,032		f_{Formato} 211,581; p. 0,000	
	$f_{\text{Interacción}}$ 27,104; p. 0,00		$f_{\text{Interacción}}$ 0,589; p. 0,440		$f_{\text{Interacción}}$ 62,996; p. 0,000	

Fuente: elaboración propia.

Los coeficientes del modelo de la izquierda de la tabla 4, interpretados en términos logarítmicos (*odds ratio*), indican que el número estimado de preguntas con efecto de primacía en el orden favorable es un 0,819 del estimado en el orden desfavorable. El resto de variables se mantienen constantes. El efecto del formato de administración sigue en esta tónica: el número estimado de preguntas con primacía cuando se utiliza el formato convencional es 0,72 veces menor que en el formato en dos preguntas. Es decir, aumenta el efecto de primacía en el orden desfavorable y en el formato en dos preguntas, así como en los niveles de estudios más altos. La interacción de las dos situaciones experimentales indica que el efecto de primacía disminuye notablemente en la interacción de las dos situaciones experimentales: orden favorable y formato convencional.

La regresión, considerando el efecto de recencia en la penúltima categoría como dependiente (parte derecha de la tabla 4), indica de nuevo la gran influencia de las dos situaciones experimentales: mayor efecto de recencia en el cuestionario que comienza por las opciones favorables y en el formato de administración convencional, en que esta última influye más que el orden de respuestas. La interacción de las dos situaciones experimentales presenta el coeficiente más elevado de todo el modelo, lo que indica que los entrevistados que comienzan a responder escuchando primero las categorías de respuestas favorables en el formato de administración convencional tienen un mayor efecto en la elección de la penúltima categoría de respuesta. Los entrevistados con estudios superiores son los que presentan mayores efectos.

Sintetizando, la encuesta telefónica empleada presenta efectos de primacía y de recencia. Los primeros afectan a ambas categorías, y los segundos, únicamente a la penúltima categoría de respuesta, con un efecto diferente al planteado en la literatura internacional sobre el tema. Los efectos de primacía son mayores en el orden desfavorable y en el formato de administración en dos preguntas, así como en las personas con mayores niveles de estudios. Los efectos de recencia en la penúltima categoría son más frecuentes en el orden favorable y en el formato convencional. Este hallazgo, teniendo en cuenta que supone la situación habitual de realización de las encuestas telefónicas,

Tabla 4. Modelo de regresión (binomial negativa) de la influencia de las situaciones experimentales y las variables sociodemográficas en los efectos de primacía y de recencia.

	V. D.: nº de respuestas efecto de primacía (ambas categorías)			V. D.: nº de respuestas efecto de recencia (penúltima categoría)		
	Coefficientes	Error estándar	Odds ratio	Coefficientes	Error estándar	Odds ratio
Intersección	2,250 ***	0,774	9,491	0,777 ***	0,118	2,175
Situación experimental						
Orden: favorable (Ref.: inverso)	-0,200 ***	0,038	0,819	0,750 ***	0,063	2,118
Formato administración: convencional (Ref.: dos preguntas)	-0,334 ***	0,040	0,716	1,019 ***	0,061	2,772
Var. sociodemográficas						
Sexo: varones (Ref.: mujeres)	0,141 ***	0,029	1,152	0,049	0,036	1,050
Edad						
Entre 16 y 30 años	-0,074	0,097	0,928	0,016	0,139	1,016
De 31 a 50 años	-0,102	0,080	0,903	-0,077	0,116	0,926
De 51 a 64 años	-0,150	0,095	0,961	0,172	0,128	1,187
Estudios terminados						
Sin estudios y primarios	-0,318 ***	0,086	0,728	-0,179	0,120	0,836
ESO	-0,178 *	0,107	0,837	-0,259 *	0,153	0,772
Bach. + FP (Ref.: superiores)	-0,241 **	0,102	0,786	-0,090	0,139	0,914
Interacciones						
Orden y formato	-0,300 ***	0,057	0,741	0,808 ***	0,077	2,243
Estud. prim. y 16-30 años	-0,297	0,280	0,743	-0,630	0,598	0,533
Estud. prim. y 35-50 años	0,133	0,116	1,142	0,049	0,155	1,051
Estud. prim. y 51-64 años	0,089	0,128	1,093	-0,266	0,166	0,767
Estud. ESO y 16-30 años	0,109	0,151	1,116	0,008	0,213	1,008
Estud. ESO y 35-50 años	-0,094	0,135	0,910	0,184	0,181	1,202
Estud. ESO y 51-64 años	-0,041	0,151	0,960	-0,178	0,200	0,837
Estud. Bach. + FP y 16-30 años	0,204	0,127	1,226	-0,035	0,174	0,965
Estud. Bach. + FP y 31-50 años	0,112	0,113	1,119	0,186	0,152	1,204
Estud. Bach. + FP y 51-60 años	0,264	0,127	1,302	-0,137	0,166	0,872
2LL _{Nulo}	1151,7			1241,4		
2LL _{Propuesto}	1005,1			823,6		
Chi Razón verosimilitud	146,6			417,8		
Pseudo R ²	0,127			0,337		
Número de casos	738			739		

* Relación significativa al 0,10.

** Relación significativa al 0,05.

*** Relación significativa al 0,01.

Fuente: elaboración propia.

Tabla 5. Tiempo necesario para responder el cuestionario (en minutos).

Orden	Formato de administración de la pregunta		Total
	Formato convencional (Cuestionario B1)	Formato con dos preguntas (Cuestionario B2)	
Favorable: cuestionario A1, orden A-B-C-D	12,17	11,91	12,04
Desfavorable: cuestionario A2, orden D-C-B-A	11,82	12,72	12,30
Total	11,99	12,35	12,17

Fuente: elaboración propia.

revela un efecto provocado no tanto por la administración oral —propia de la encuesta telefónica—, sino por efectos relacionados con el cuestionario (como son el orden de exposición de los estímulos y el formato de administración de la pregunta).

Quedaría por contrastar la hipótesis referida al tiempo necesario para responder los cuestionarios, considerando que el mayor número de preguntas de los cuestionarios B necesitará más tiempo. En la tabla 5, se muestra el tiempo utilizado para responder el cuestionario, y destaca una mayor duración en los cuestionarios que comienzan con la opción desfavorable (segunda fila, 12,30 minutos). Aunque la duración aumenta todavía más en el formato convencional, son diferencias pequeñas que no llegan a ser significativas, lo que implica no poder aceptar la cuarta hipótesis. La única diferencia significativa es la comparación entre la duración del cuestionario favorable en formato convencional y el cuestionario desfavorable en dos preguntas, 12,17 y 12,72 minutos, respectivamente.

4. Conclusiones

Este trabajo surge tras la lectura de la sección «Diseño óptimo del cuestionario», incluida dentro de las conclusiones de un artículo de Holbrook et al. (2007) sobre los efectos de respuesta en preguntas dicotómicas utilizando 548 investigaciones realizadas por Gallup en los Estados Unidos. En las dos últimas páginas recomienda, en primer lugar, rotar el orden de respuesta de las preguntas con el fin de evitar sesgos en los resultados. Los efectos de respuesta, según su experiencia, son mayores en preguntas difíciles de comprender, por lo que sugiere a los investigadores utilizar un lenguaje lo más sencillo posible (segunda recomendación). En tercer lugar, aconseja el uso de cuestionarios breves, ya que su investigación detecta que los efectos de respuesta aumentan a medida que transcurre la entrevista.

Se ha tratado de considerar lo acertado de la primera recomendación (rotar) cumpliendo la segunda y la tercera, ya que la presente investigación emplea un lenguaje sencillo, con preguntas cortas y pocas opciones de respuesta, que dan lugar a utilizar un cuestionario para responder en alrededor de 13 minutos.

El contraste de la primera hipótesis, que proponía una mayor presencia de efectos de recencia en la encuesta telefónica, desvela lo acertado de su formulación, ya que el 22,5 % de las preguntas del cuestionario han experimentado este efecto (tabla 1), si bien es un efecto que convive con la primacía del 21,1 % de las preguntas. Son diferencias pequeñas (y no significativas) que llevan a cuestionar la aplicabilidad de lo detectado por otras investigaciones a la situación en nuestro país. En cualquier caso, se produce una situación muy diferente a la localizada en las encuestas autoadministradas, donde presentar primero las opciones de respuesta favorables suponía respuestas más rápidas (entre otros, Tourangeau et al., 2013b; Dillman et al., 2014).

Las escasas diferencias entre la presencia de efectos de primacía y de recencia —señaladas en el párrafo anterior— aumentan notablemente cuando se cambia el orden de administración de las categorías y el formato de administración (tabla 2). Respecto al primero —el efecto del orden—, la recencia suave (penúltima categoría) es la única que aumenta cuando se utiliza un orden favorable, y se localiza también un efecto de primacía suave en el orden desfavorable. Ahora bien, el efecto de recencia es mayor que el de primacía.

Resulta curioso, por otra parte, que los efectos de primacía y de recencia se concentren en las categorías centrales, es decir, en la segunda y en la penúltima, y no en las más extremas.

El formato de administración presenta un elevado efecto de primacía en el formato en dos preguntas, mientras que el efecto de recencia funciona de forma diferente según se considere la última o la penúltima categoría. En cualquier caso, utilizar un formato en dos preguntas muestra un efecto de extremidad derivado de la reducción del número de categorías, en línea con lo que se propuso en la segunda hipótesis. Este efecto, que plantea una mayor elección de las categorías más extremas, tiene lugar en tres de las cuatro situaciones analizadas: primacía extrema, primacía suave y recencia penúltima. Debe tenerse en cuenta que la influencia del formato de administración es superior al localizado en el orden de presentación de las respuestas (tabla 2).

La gran influencia del formato de administración, sin embargo, no llega a eliminar los efectos derivados del orden en el que se presentan las categorías de respuesta (tercera hipótesis). Solo en el caso de recencia extrema tiene lugar esta situación, y desaparecen las diferencias en el orden de administración de las categorías (tablas 3 y 4).

Por otro lado, el aumento del número de preguntas que supone la descomposición de preguntas de cuatro o cinco categorías en dos preguntas apenas hace variar el tiempo necesario para responder el cuestionario (cuarta hipótesis), lo que mantiene sin cambios el coste del trabajo de campo (tabla 5).

El estudio de las variables sociodemográficas indica que las personas con mayores niveles de estudios presentan los mayores efectos de respuesta, pero no existe diferencia significativa según la edad. Los hombres están más influidos que las mujeres en los efectos aquí analizados.

Diversos trabajos (entre otros, Krosnick, 1999; Galesic et al., 2008; Hollbrook et al., 2007) destacan que los efectos de respuesta se incremen-

Tabla 6. Efecto de primacía y de recencia en investigaciones realizadas en la Comunidad Autónoma Vasca con encuestas telefónicas.

De las siguientes instituciones o colectivos, ¿qué tres cree usted que tienen más poder en el País Vasco? (Leer todas las instituciones o colectivos) (Máximo tres respuestas)

ORDEN; Parlamento, Gobierno, partidos, bancos...

ORDEN: empresas, militares, sindicatos, medios, banco...

Las grandes empresas	47 %	34 %
Los militares	6 %	3 %
Los sindicatos	12 %	12 %
Los medios de comunicación	35 %	30 %
Los bancos	37 %	32 %
Los partidos políticos	25 %	37 %
El Gobierno Vasco	56 %	56 %
El Parlamento vasco	24 %	29 %

Nota: se han colocado en negrilla las cifras con diferencias de 5 puntos porcentuales.

Fuente: Gabinete de Prospección Sociológica (Gobierno Vasco). Reproducido de Martínez de Luna, 2008: 14.

tan cuando aumenta el número de información, y es posible que no aparezcan aquí tan nítidamente por el esfuerzo puesto en la elaboración del cuestionario, al reducir al máximo la información, limitar el número de palabras, llevar a cabo dos pretest y realizar un detallado proceso de formación y supervisión del personal de campo. Se recomienda precaución a la hora de generalizar estos resultados, porque son conclusiones extraídas de un único estudio realizado en una comunidad. Aunque estos resultados debieran contrastarse con investigaciones de otras temáticas realizadas en otras zonas geográficas, uno de los escasos estudios disponibles en nuestro país —realizado en la Comunidad Autónoma Vasca— desvela aumentos del efecto de recencia en preguntas con ocho categorías de respuesta. Tal y como se muestra en la tabla 6, las grandes empresas, los medios de comunicación y los bancos son más elegidos cuando son leídos al final de la lista. También sucede lo mismo con los partidos políticos y el Gobierno Vasco. Las diferencias rondan los 5 puntos porcentuales, y alcanzan los 13 en el caso de las grandes empresas.

La mejor solución, sin duda, es la aleatorización de las opciones de respuesta, algo tremendamente sencillo en las encuestas electrónicas (es como son la mayor parte de las encuestas telefónicas implantadas en sistemas CATI), aunque no siempre utilizado. Esto mejora, sin duda, las preguntas nominales con mucha información, pero no es aplicable al nivel de medición ordinal, que siempre debe seguir una secuencia lógica de más a menos o de menos a más. Solo la categoría intermedia (o regular) admitiría cambiar de lugar, aunque algunos trabajos (Tourangeau et al., 2013b) detectan un aumento del tiempo para responder cuando no está colocada en el centro. Ahora bien, también presentan un menor número de elecciones cuando se coloca al final.

Referencias bibliográficas

- AZOFRA, María José (1999). *Cuestionarios*. Colección «Cuadernos Metodológicos», 26. Madrid: Centro de Investigaciones Sociológicas.
- BISHOP, George; SMITH, Andrew (1997). *Response-order effects in public opinion surveys: The plausibility of rival hypotheses*. Presentado en la conferencia anual de la American Association for Public Opinion Research, Norfolk, VA.
- (2001). «Response-Order Effects and the Early Gallup Split-Ballots». *Public Opinion Quarterly*, 65, 479-505.
<<https://doi.org/10.1086/323575>>.
- CANNEL, Charles (1992). *Characteristics of questions that cause problems for interviewer and respondents*. Ann Arbor: Survey Research Center.
- CANTRIL, Hadley (1944). *Gauging Public Opinion*. Princeton: Princeton University Press.
- CHANG, Linchat; KROSNICK, Jon A. (2010). «Comparing oral interviewing with self-administered computerized questionnaires: An experiment». *Public Opinion Quarterly*, 74, 154-167.
<<https://doi.org/10.1093/poq/nfp090>>.
- CHRISTIAN, Leah Melani; DILLMAN, Don A.; SMYTH, Jolene D. (2007). «Helping Respondents Get it Right the First Time: The Relative Influence of Words, Symbols, and Graphics in Web and Telephone Surveys». *Public Opinion Quarterly*, 71 (1), 113-125.
<<https://doi.org/10.1093/poq/nfl039>>.
- (2008). «The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys». En: LEPKOWSKI, James; et al. (eds.). *Advances in Telephone Survey Methodology*. Nueva York: Wiley-Interscience.
- CHRISTIAN, Leah Melani; SMYTH, Jolene D. (2008). «Visual Communication». En: LAVRAKAS, P. (ed.). *Encyclopedia of Survey Research Methods*. California: Sage Publications, Inc.
- CONRAD, Frederick (2019). «Race of Virtual Interviewer Effects». *Interviewers and their effects from a total survey error perspective*. International Workshop celebrado en la Universidad de Nebraska-Lincoln, 26-28 de febrero.
- DÍAZ DE RADA, Vidal (2011). «Utilización conjunta de encuesta presencial y telefónica en encuestas electorales». *Revista Internacional de Sociología*, 69, 393-416.
<<https://doi.org/10.3989/ris.2009.01.10>>.
- DÍAZ DE RADA, Vidal; AYERDI, Peio (2007). «Algunos problemas de la encuesta telefónica para la proyección electoral». *Revista Española de Investigaciones Sociológicas*, 118, 153-204.
- DÍAZ DE RADA, Vidal; PORTILLA, Idoia (2015). «Encuestas telefónicas: estrategias para mejorar la colaboración». *Perspectiva Empresarial*, 2 (1), 97-115.
<<https://doi.org/10.16967/rpe.47>>.
- DILLMAN, Don A. (1978). *Mail and Telephone Surveys*. Nueva York: Wiley.
- DILLMAN, Don A.; GERTSEVA, Arina; MAHON-HAFT, Taj (2005). «Achieving usability in establishment surveys through the application of visual design principles». *Journal of Official Statistics*, 21 (2), 183-214.
- DILLMAN, Don A.; SMYTH, Jolene D. (2007). «Design Effects in the Transition from Telephone to Web-Based Surveys». *American Journal of Preventive Medicine*, 32 (5S), S90-S96.
<<https://doi.org/10.1016/j.amepre.2007.03.008>>.

- DILLMAN, Don A.; SMYTH, Jolene D.; CHRISTIAN, Leah Melanie (2014). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. 4ª edición. Nueva York: John Wiley.
- ELLIS, Charles H.; KROSNICK, Jon (1999). *Comparing telephone and face to face surveys in terms of sample representativeness: A Meta-Analysis of Demographics Characteristics*. Ann Arbor: Universidad de Michigan, NES.
- FOWLER, Floyd J. (1995). *Improving Survey Questions*. Londres: Sage.
- (2011). «Coding the behavior of interviewers and respondents to evaluate survey questions». En: MADANS, Jennifer; MILLER, Kristen; MATTLAND, Aaron; WILLIS, Gordon (2011). *Question Evaluation Methods: Contributing to the Science of Data Quality*, 7-22. Nueva York: Wiley.
- FREY, James H. (1989). *Survey Research by Telephone*. Londres: Sage.
- GALESIC, Mirta; TOURANGEAU, Robert; COUPER, Mick P.; CONRAD, Frederick (2008). «Eye-tracking data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding». *Public Opinion Quarterly*, 72 (5), 892-913. <<https://doi.org/10.1093/poq/nfn059>>.
- GARCÍA FERRANDO, Manuel; LLOPIS, Ramón (2015). «La encuesta». En: GARCÍA FERRANDO, Manuel; et al. (comps.). *El análisis de la realidad social*. Madrid: Alianza.
- GILBERT, Emily E. (2015). «A comparison of branched versus unbranched rating scales for the measurement of attitudes in surveys». *Public Opinion Quarterly*, 79, 443-470. <<https://doi.org/10.1093/poq/nfu090>>.
- GROVES, Robert M. (2012). «Three Eras of Survey Research». *Public Opinion Quarterly*, 76, 861-871. <<https://doi.org/10.1093/poq/nfr057>>.
- GROVES, Robert M.; KAHN, Louis (1979). *Surveys by telephone: a national comparison with personal interviews*. Orlando: Academic Press.
- GWARTNEY, Patricia A. (2007). *The Telephone Interviewer's Handbook: How to conduct standardized conversations*. San Francisco: Jossey-Bass.
- HÄDER, Sabine; HÄDER, Michael; KÜHNE, Mike (2012). «Introduction: telephone surveys in Europe». En: HÄDER, Sabine; HÄDER, Michael; KÜHNE, Mike (eds.). *Telephone Surveys in Europe: Research and Practice*. Mannheim: Springer Verlag.
- HOLBROOK, Allyson L.; GREEN, Melanie C.; KROSNICK, Jon A. (2003). «Telephone versus face-to-face interviewing of national probability samples with long questionnaires». *Public Opinion Quarterly*, 67, 79-125. <<https://doi.org/10.1086/346010>>.
- HOLBROOK, Alyson L.; KROSNICK, Jon A.; MOORE, David; TOURANGEAU, Robert (2007). «Response order effects in dichotomous categorical questions presented orally: the impact of question and respondent attributes». *Public Opinion Quarterly*, 71, 325-348. <<https://doi.org/10.1093/poq/nfm024>>.
- JOHNSON, Timothy P.; HOLBROOK, Allyson; CHO, Young Ik; SHAVITT, Sharon; CHAVEZ, Noel; WEINER, Saur (2019). «Examining the Comparability of Behavior Coding Across Cultures». En: JOHNSON, Timothy P.; PENNEL, Beth-Ellen; STOOP, Ineke; DORER, Brita, (eds), *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*. Hoboken, N.J.: Wiley.
- KROSNICK, Jon A. (1991). «Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys». *Applied Cognitive Psychology*, 5, 213-236. <<https://doi.org/10.1146/annurev.psych.50.1.537>>.
- (1999). «Survey research». *Annual Review of Psychology*, 50, 537-567. <<https://doi.org/10.1146/annurev.psych.50.1.537>>.

- KROSINICK, Jon A.; ALWIN, Duane F. (1987). «An evaluation of a cognitive theory of response-order effects in survey measurement». *Public Opinion Quarterly*, 51, 201-219.
<<https://doi.org/10.1086/269029>>.
- KROSINICK, Jon A.; BERENT, Katthew M. (1993). «Comparison of party identification and policy preferences: the impact of survey question format». *American Journal of Political Science*, 37, 941-964.
<<https://doi.org/10.2307/2111580>>.
- KROSINICK, Jon A.; NARAYAN, Sowmya; SMITH, Wendy R. (1996). «Satisficing in Surveys: Initial Evidence». *New Directions for Evaluation*, 70, 29-44.
<<https://doi.org/10.1002/ev.1033>>.
- KROSINICK, Jon A.; HOLBROOK, Allyson L.; BERENT, Matthew K.; et al. (2002). «The Impact of “No Opinion” Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?». *Public Opinion Quarterly*, 66, 371-403.
<<https://doi.org/10.1086/341394>>.
- KROSINICK, Jon A.; PRESSER, Stanley (2010). «Question and questionnaire design». En: MARSDEN, Peter; WRIGHT, J. (eds.). *Handbook of Survey Research*. Bingley: Emerald Group Publishing Limited.
- LAVRAKAS, Paul J. (1993). *Telephone Survey Methods*. Newbury Park-California: Sage.
- (2010). «Telephone survey». En: MARSDEN, Peter V.; WRIGHT, James D. (eds.). *Handbook of Survey Research*. Bingley: Emerald Group Publishing Limited.
- LEEuw, Edith de (2008). «Choosing the method of data collection». En: LEEuw, Edith de; HOX, Jon J.; DILLMAN, Don A. (eds.). *International Handbook of Survey Methodology*. Nueva York: Lawrence Erlbaum Associates and European Association of Methodology.
- LEEuw, Edith de; HOX, Joop J. (2015). «Survey mode and mode effects». En: ENGEL, Uwe; et al. (eds.). *Improving survey methods. Lessons from recent research*. Nueva York y Londres: Routledge and European Association of Methodology.
- LUNDMARK, Sebastian; GILLJAM, Mikael; DAHLBERG, Stefan (2016). «Measuring generalized trust: An examination of question wording and the number of scale points». *Public Opinion Quarterly*, 80 (1), 26-43.
<<https://doi.org/10.1093/poq/nfv042>>.
- MALHOTRA, Neil; KROSINICK, Jon A.; THOMAS, Randall, K. (2009). «Optimal design of branching questions to measure bipolar construct». *Public Opinion Quarterly*, 73, 304-324.
<<https://doi.org/10.1093/poq/nfp023>>.
- MARTÍNEZ DE LUNA, Iñaki (2008). «Encuestas de opinión: de la teoría a la práctica». *Metodología de Encuestas*, 10, 7-26.
- MAYNARD, Douglas W.; SHAEFFER, Nora Cate (2006). «Standardization in Interaction: The Survey Interview». En: DREW, Paul; RAYMOND, Geoff; WEINBERG, Darin (eds.). *Talking Research*. Londres: Sage.
- NICOLAAS, G.; THOMSON, K.; LYNN, P. (2000). *The feasibility of conducting Electoral Surveys in the UK by telephone*. Londres: National Centre for Social Research.
- OLSON, Kristen; SMYTH, Jolene D. (2015). «The Effect of CATI Questions, Respondents, and Interviewers on Response Time». *Journal of Survey Statistics and Methodology*, 3 (3), 361-396.
<<https://doi.org/10.1093/jssam/smv006>>
- PARK, Hyunjoo; LEE, Jaehwan (2018): «Exploring the Validity of Behavior Coding». *Field Methods*, 30 (3), 225-240.
<<https://doi.org/10.1177/1525822X18781881>>.

- PAYNE, Stanley L. (1951). *The art of asking questions*. Princeton: Princeton University Press.
- ROSSI, Peter H.; WRIGHT, James D.; ANDERSON, Andy B. (1983). «Sample surveys: history, current practice and future prospects». En: ROSSI, Peter H.; WRIGHT, James D.; ANDERSON, Andy B. (eds.). *Handbook of Survey Research*. Nueva York: Academic Press.
- SCHWARZ, Norbert; HIPPLER, Hans J. (1987). «What response scale may tell your respondents: Information functions of response alternatives». En: HIPPLER, Hebert J.; SCHWARZ, Norbert; SUDMAN, Seymour (eds.). *Social information processing and survey methodology*. Nueva York: Springer-Verlag.
- SINGER, Eleanor (2016). «Reflections on Surveys' Past and Future». *Journal of Survey Statistics and Methodology*, 4, 463-475.
<<https://doi.org/10.1093/jssam/smw026>>.
- SMYTH, Jolene D.; DILLMAN, Don A.; CHRISTIAN, Leah Melani (2007). «Context Effects in Web Surveys: New Issues and Evidence». En: JOINSON, Alain; MCKENNA, Katelyn; POSTMES, Tom. Nueva York: Oxford University Press.
- STERN, Michael J.; SMYTH, Jolene D.; MENDEZ, J. (2012). «The Effects of Item Salience and Question Design on Measurement Error in a Self-Administered Survey». *Field Methods*, 24, 3-27.
<<https://doi.org/10.1177/1525822X11419478>>.
- TOURANGEAU, Roger; RASINSKI, Kennet (1988). «Cognitive Processes Underlying Context Effects in Attitude Measurement». *Psychological Bulletin*, 103, 299-314.
<<https://dx.doi.org/10.1037/0033-2909.103.3.299>>.
- TOURANGEAU, Roger; COUPER, Mick P.; CONRAD, Frederick (2004). «Spacing, Position, and Order: Interpretive heuristics for visual features of survey questions». *Public Opinion Quarterly*, 68, 368-393.
<<https://doi.org/10.1093/poq/nfh035>>.
- (2013a). «Up Means Good»: The Effect of Screen Position on Evaluative Ratings in Web Surveys». *Public Opinion Quarterly*, 77, 69-88.
<<https://doi.org/10.1093/poq/nfh035>>.
- (2013b). *The Science of Web Surveys*. Nueva York: Oxford University Press.
- VISAUTA, Restituto (1989). *Técnicas de investigación social*. Barcelona: PPU.
- WRIGHT, James D.; MARDSEN, Peter V. (2010). «Survey research and social science: history, current practice and future prospects». En: MARDSEN, Peter V.; WRIGHT, James D. (eds.). *Handbook of Survey Research*. Bingley: Emerald Group Publishing Limited.
- YE, Cong; FULTON, Henna; TOURANGEAU, Roger (2011). «More positive o more extreme? A meta-analysis of mode difference in response choice». *Public Opinion Quarterly*, 75, 349-3.655.
<<https://doi.org/10.1093/poq/nfr009>>.
- ZALLER, John (2014). *La naturaleza y los orígenes de la opinión pública*. Madrid: Centro de Investigaciones Sociológicas.
- ZHANG, Chan (2013). «Design of Interactive Interventions in Web Surveys: Humanness, Social Presence, and Data Quality». *Satisficing in Web Surveys: Implications for Data Quality and Strategies for Reduction. Doctor of Philosophy (Survey Methodology)*, 60-96.
- ZOUDEN, J. van der; SMITH, J. H. (2004). «Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach». En: PRESSER, S.; ROTHGEB, J.; COUPER, M.; Lessler, J.; Martin, E.; Martin, J.; Singer, E. (eds.). *Questionnaire Development Evaluation and Testing Methods*, 109-130. Nueva York: Wiley.

Anexo 1. Temática de las preguntas de la encuesta telefónica en una comunidad autónoma española

Temática (nº de variables)	Amplitud de la respuesta
Economía y política de la comunidad. Economía personal y principales problemas	
Valoración de la situación económica de la comunidad	1-5
Valoración de la situación económica personal	1-5
Valoración de la situación política de la comunidad	1-5
Política	
Interés en la política	1-4
Frecuencia de seguimiento de la política a través de diferentes fuentes de información (3)	1-5
Política de la comunidad	
Valoración de los portavoces parlamentarios conocidos (7)	0-10
Escala de autoubicación ideológica izquierda-derecha	0-10
Sentimiento europeísta personal	1-4
Satisfacción con el nivel de autonomía de la comunidad	1-5*
Deseo de aumentar o disminuir el nivel de autonomía	1-4
El Parlamento de la comunidad y sus órganos	
Valoración de la actividad del Defensor del Pueblo de la comunidad	0-10
Valoración de la actividad del Defensor de la Cámara de Comptos	0-10
Valoración de la actividad del Parlamento de la comunidad	0-10
Valoración de la actuación política de los diferentes partidos políticos	1-5*
Gobierno autonómico y aspectos diversos	
Nivel de acuerdo con que el Gobierno autonómico lleve a cabo una cooperación permanente con una comunidad vecina	1-5*
Valoración del Gobierno en cuanto a su capacidad de resolver los problemas	1-4
Valoración de la gestión del Gobierno de la comunidad	0-10
Valoración de la gestión del presidente de la comunidad	0-10
Estimación del nivel de abstención si mañana hubiera elecciones al Parlamento autonómico	0-10

* Nota: son preguntas con cinco categorías de respuesta en las que solo cuatro son leídas al entrevistado, sin leer la opción intermedia.

Anexo 2. Características de las cuatro muestras empleadas en la encuesta telefónica (porcentajes verticales)

	Tipo de cuestionario			
	Orden de respuestas		Formato de administración	
	Favorable	Desfavorable	Convencional	Dos preguntas
	(A1)	(A2)	(B1)	(B2)
Sexo (Ji-Cuadrado NO significativo)				
Hombre	51,9 %	50,5 %	49,8 %	51,0 %
Mujer	48,1 %	49,5 %	50,2 %	49,0 %
Edad (Ji-Cuadrado NO significativo)				
De 16 a 30 años	16,5 %	17,3 %	17,1 %	17,0 %
De 31 a 50 años	38,2 %	38,3 %	41,5 %	39,2 %
De 51 a 64 años	22,8 %	21,3 %	21,0 %	21,8 %
Más de 64 años	22,5 %	23,1 %	20,5 %	22,2 %
Nivel de estudios (Ji-Cuadrado NO significativo)				
Sin estudios y primarios	19,7 %	23,2 %	19,2 %	20,8 %
ESO	14,3 %	10,7 %	14,3 %	13,0 %
Bachillerato y FP	37,6 %	41,2 %	38,4 %	39,1 %
Superiores	28,3 %	25,0 %	28,1 %	27,1 %
Relación con la actividad (Ji-Cuadrado NO significativo)				
Ocupado/a	49,1 %	45,3 %	49,5 %	47,8 %
Parado/a	6,5 %	8,0 %	9,4 %	7,9 %
Jubilado/a	26,5 %	27,7 %	22,8 %	26,0 %
Estudiante	12,0 %	13,1 %	11,9 %	12,4 %
Trabajo doméstico no remunerado	5,8 %	5,8 %	6,4 %	6,0 %
Nº de casos	200	200	200	20