

ESCALAS EN PREGUNTAS DE BATERÍA EN ENCUESTAS TELEFÓNICAS: ¿CAMBIAN LAS RESPUESTAS CUANDO LAS PREGUNTAS SE PLANTEAN DE FORMA INDEPENDIENTE?

VIDAL DÍAZ DE RADA

Departamento de Sociología y Trabajo Social, I-COMMUNITAS. Institute for Advanced Social Research, Universidad Pública de Navarra, Pamplona.
email: vidal@unavarra.es
ORCID iD: <http://orcid.org/0000-0002-9638-3741>

Cómo citar este artículo / Citation: Díaz de Rada, V. 2023. Escalas en preguntas de batería en encuestas telefónicas: ¿cambian las respuestas cuando las preguntas se plantean de forma independiente?, *Revista Internacional de Sociología* 81 (4): e240. <https://doi.org/10.3989/ris.2023.81.4.M23-2>

RESUMEN

Este trabajo versa sobre las denominadas ‘preguntas de batería’, un conjunto de preguntas *encadenadas*, que se complementan entre sí al estar referidas a una misma temática y utilizar la misma respuesta. Estas preguntas, muy habituales en los cuestionarios, presentan ventajas e inconvenientes frente a la descomposición de los diferentes elementos en preguntas independientes. El presente trabajo utiliza una encuesta telefónica a población general para comprobar hasta qué punto las preguntas que forman parte de una batería se responden peor que cuando se trata de preguntas independientes. Se han considerado tres preguntas, una actitudinal y dos de comportamiento, todas ellas con cinco opciones de respuesta. La distribución de las preguntas, unida a la utilización de siete indicadores de calidad en la respuesta, muestra que las preguntas de batería no se responden peor que cuando las mismas preguntas son consideradas de forma independiente.

PALABRAS CLAVE

Encuesta telefónica; Preguntas Múltiples; Escala Likert; Aquiescencia; Diferenciación; Calidad en la respuesta en encuestas.

GRID QUESTIONS IN TELEPHONE SURVEYS. DO THE ANSWERS CHANGE WHEN THE QUESTIONS ARE ASKED INDEPENDENTLY?

Copyright: © 2023 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)

Recibido: 07.11.2022. **Aceptado:** 24.08.2023.
Publicado: 12.01.2024

ABSTRACT

This work deals with the so-called “grid questions”: a series of items is presented (usually in rows), sharing a common set of response options (usually in columns), asking one or more questions about each item. These questions, very common in questionnaires, have advantages and disadvantages compared to the decomposition of the different elements into independent questions. This paper uses a telephone survey of the general population to analyze the extent to which the questions that are part of a battery are answered worse than when they are independent questions. Three questions have been considered, one attitudinal and two behavioral, all of them with five response categories. The distribution of the questions, together with the use of seven quality indicators in the answer, shows that the battery questions are not answered worse than when the same questions are considered independently.

KEYWORDS

Telephone survey; Multiple question; Liker scale; Acquiescence; Straightlining; Response quality in surveys.

INTRODUCCIÓN

La principal clasificación de los tipos de preguntas en cuestionarios es la que considera la libertad de respuesta del encuestado (entre otros, Schuman 2008: 30). Esto permite diferenciar entre preguntas *abiertas* —en las que el entrevistado puede responder lo que se le antoje— y *cerradas*, que limitan la elección del encuestado a una serie de respuestas previamente formuladas por el investigador.

Las preguntas cerradas son clasificadas en función del número de respuestas a recoger, que da lugar al segundo nivel de clasificación. Las preguntas con respuesta única precisan de opciones de respuesta exhaustivas y mutuamente excluyentes que no hagan dudar al encuestado de qué respuesta elegir. Dentro de estas se diferencia cuando existe, o no, una relación de orden entre las respuestas. Cuando las opciones implican una relación de orden y se desea profundizar en un determinado tema, es habitual elaborar preguntas con la misma formulación utilizando las mismas opciones de respuesta. Son las denominadas ‘preguntas de batería’, que constituyen el objeto de estudio de este trabajo.

Un gran número de expertos (entre otros, recientemente, Terentev y Maloshonok 2019; Toepoel, Vermeeren y Metin 2019) han detectado que el diseño de las opciones de respuesta tiene un efecto significativo en las respuestas de las personas encuestadas. Partiendo de estos hallazgos, este trabajo utiliza dos muestras equivalentes para comparar las respuestas obtenidas en una pregunta de batería y en una serie de preguntas con la misma temática y que son administradas secuencialmente (sin formar parte de una batería). Se utilizan tres preguntas, una actitudinal (siete ítems) y dos sobre comportamientos (cinco y seis ítems respectivamente), todas respondidas utilizando una escala tipo Likert de cinco opciones. El objetivo es averiguar si el uso de una pregunta en batería, o “descompuesta” en varias preguntas sencillas, presenta cambios en la respuesta. Estas posibles diferencias se consideran en relación con:

- la ausencia de respuestas por no saber o declinar responder;
- la elección de respuestas extremas;
- el número de elecciones de respuestas “positivas” o favorables (aquiescencia);
- la presencia de “respuestas fáciles”, como son los extremos o el centro de la escala;
- el número de respuestas centrales;
- la correlación entre las respuestas;
- las elecciones de la misma respuesta en varias preguntas diferentes (ausencia de diferenciación); y
- la distribución de las respuestas.

La hipótesis de partida es que las preguntas que forman parte de una batería presentan con más frecuencia cada una de estas situaciones.

Para ello, se utiliza una encuesta telefónica aplicada a dos muestras equivalentes. Debe tenerse en cuenta que esta modalidad implica una comunicación estrictamente oral, sin las ayudas visuales que proporcionan las encuestas presenciales (utilizando tarjetas con las opciones de respuesta) ni las autoadministradas, que incluyen figuras y otros recursos gráficos en las opciones de respuesta (Toepoel y Dillman 2011). La ausencia de comunicación visual en la encuesta telefónica explica, a juicio de un gran número de expertos (entre otros, De Leeuw y Hox 2015; Jordan, Marcus y Reeder 1980; Groves y Kahn 1979), la peor respuesta de las preguntas de batería.

Es importante dar cuenta de que apenas existen investigaciones que consideren la influencia de este tipo de preguntas en las encuestas telefónicas, frente a la gran cantidad de investigaciones con encuestas autoadministradas a través de la web (entre otros, Revilla, Toninelli y Ochoa 2017; Kim *et al.* 2019; Liu y Cernat 2018; Revilla y Couper 2018; Revilla, Saris y Krosnick 2014; Couper *et al.* 2013; Chang y Krosnick 2010), fundamentalmente en el ámbito anglosajón. Se trata de una situación sorprendente cuando diversas investigaciones realizadas en otros contextos (Greene, Speizer y Wiitala 2008; Holbrook, Green y Krosnick 2003) han detectado una peor respuesta por parte de los cuestionarios telefónicos.

El artículo comienza con una revisión de los trabajos publicados sobre el tema, para proceder a continuación con la metodología, cuando se da cuenta del diseño de la investigación, El muestreo, las medidas y los análisis de datos. Los resultados preceden a las conclusiones y principales hallazgos. Un listado con las referencias bibliográficas y tres anexos cierran el trabajo.

DEFINICIÓN CONCEPTUAL

Se definen como preguntas de batería una serie de preguntas “ligadas” que se complementan entre sí con el fin de profundizar en los aspectos objeto de estudio. Se trata de cuestiones que utilizan la misma formulación para preguntar sobre aspectos diferentes utilizando las mismas opciones de respuesta (Olson, Smyth y Cochran 2018; Azofra 1999). Otros expertos las definen como “un conjunto de preguntas *encadenadas*, que se complementan entre sí al estar referidas a una misma temática y utilizar la misma respuesta” (Díaz de Rada 2001: 87). Más precisa es la definición de Couper *et al.* (2013: 323): “en estas preguntas se presentan varios elementos (generalmente en filas), compartiendo

un conjunto común de opciones de respuesta (generalmente en columnas), haciendo una o más preguntas sobre cada elemento”. Reciben este nombre por la forma que tienen en el cuestionario, similar a un cuadro de doble entrada.

Son muy frecuentes en los cuestionarios y, de hecho, un análisis de los barómetros del CIS del primer semestre del año 2022 desvela que más de la mitad de las respuestas de las preguntas de “contenido” (excluidas las sociodemográficas) son preguntas de batería (véase anexo 1). Su elevado nivel de utilización se justifica porque dan más fluidez al cuestionario y su capacidad para profundizar en las temáticas objeto de estudio.

Esta gran utilización, a juicio de Lozar Manfreda, Vehovar y Koren (2008), se explica por su sencillez de preparación, la posibilidad de medir un gran número de conductas, así como una administración del cuestionario fácil y rápida (al tener que ser explicadas una vez). Además, presentan un coste menor que el resto de preguntas: menos coste de papel, menos trabajo de los entrevistadores y economía en la preparación (Revilla, Saris y Krosnick 2014). Otros expertos destacan que el entrevistado responde entre dos y tres veces más rápido que en otras preguntas (Dillman, Smyth y Christian 2014; Bethlehem y Biffignandi 2011; Couper, Traugott y Lamias 2001). En definitiva, eficientes y fáciles de administrar (Kim et al. 2019).

La suposición subyacente de estas preguntas es que las personas encuestadas diferenciarán las respuestas a los diferentes ítems, pero las baterías no siempre producen datos de alta calidad (Couper et al. 2013). Bien es verdad que su abuso puede cansar al entrevistado y perjudicar la calidad de la información por el riesgo de elegir —de forma automática y sin pensar— la misma respuesta; situación definida como ausencia de *diferenciación* (Kim et al. 2019; Krosnick y Alwin 1988).

Otro problema de las preguntas de batería es que, al presentar varias veces la misma respuesta, los entrevistados seleccionen las respuestas extremas, algo habitual en las preguntas administradas sin ayudas visuales, como sucede en las encuestas telefónicas, como la utilizada en este trabajo (De Leeuw y Hox 2015; Ye, Fulton y Tourangeau 2011; Nicolaas, Thomson y Lynn, 2000; De Leeuw, 1992; Sykes y Collins 1988; Jordan, Marcus y Reeder 1980; Groves y Kahn 1979). Esta elección de los extremos supone, la mayor parte de las veces, una mayor elección de respuestas “positivas” o favorables (Ye, Fulton y Tourangeau 2011; Sykes y Collins 1988), conocida como *aquiescencia*. La “pesadez” de una pregunta en que se proporciona varias veces la misma respuesta lleva, en ocasiones, a la elección de respuestas “fáciles”, como es la elección de los extremos o el centro de la escala utilizada.

Otra desventaja, señalan Revilla, Toninelli y Ochoa (2017), es el aumento de la correlación entre ítems que miden el mismo concepto, así como una menor satisfacción de las personas entrevistadas, pese al ahorro de tiempo que supone su utilización. Otras investigaciones, como la realizada por Liu y Cernat (2018), detectan una escasa equivalencia entre baterías y preguntas separadas cuando hay más de siete alternativas de respuesta.

En el cuadro 1, se muestra una síntesis de las ventajas y desventajas de este tipo de preguntas, incluidas las relacionadas con el modo autoadministrado, para el que no son muy adecuadas porque suelen presentar un elevado número de preguntas sin responder y un notable incremento del número de abandonos en la cumplimentación del cuestionario (Liu 2017; Couper et al. 2013; Peytchev 2011). De hecho, diversas investigaciones recomiendan limitar, e incluso suprimir, este tipo de preguntas por su gran dificultad para ser respondidas (Revilla, Toninelli y Ochoa 2017; Dillman, Smyth y Christian 2014; Gräf 2002; Wojtowicz 2001). A pesar de estos límites siguen siendo muy utilizadas en todo tipo de cuestionarios, tanto aplicados de forma presencial, telefónica y autoadministrada (véase el anexo 1).

Cuadro 1.
Ventajas y desventajas de las preguntas de batería

Ventajas
• sencillez de preparación,
• posibilidad de medir un gran número de conductas,
• administración del cuestionario fácil y rápida (solo se explican una vez),
• menor coste que el resto de preguntas:
» menos coste de papel,
» menos trabajo de los entrevistadores,
» economía en la preparación,
• rapidez de respuesta.
Desventajas
• mayor cansancio del entrevistado,
• menor calidad de la información debido a:
» repetición de respuestas en todos los ítems (ausencia de diferenciación)
» selección de extremas,
» mayor aquiescencia,
» elección de respuestas fáciles,
» aumento de la correlación entre ítems,
• menor satisfacción de los encuestados con la encuesta,
• en el modo autoadministrado:
» elevado número de preguntas sin responder,
» incremento del número de abandonos sin terminar de responder.

Fuente: elaboración propia con base en las referencias citadas

METODOLOGÍA

Diseño

Para comprobar hasta qué punto el formato de pregunta utilizado (considerando baterías o preguntas sueltas) influye en las respuestas, se elaboraron dos cuestionarios idénticos, excepto en tres preguntas, una actitudinal y dos referidas a comportamientos. La primera considera el nivel de satisfacción con diversos aspectos de la vida de la persona encuestada (salud, vida familiar, relaciones de pareja, amistades, situación laboral, tiempo libre disponible y distribución de las tareas del hogar); la segunda se refiere a la presencia de problemas de salud (dificultades en el trabajo, achaques o dolores, sensación de desdicha o depresión, pérdida de confianza en sí mismo, sensación de no ser capaz de superar sus problemas). Ambas se responden utilizando una escala tipo Likert de cinco opciones. La tercera pregunta analiza el reparto de seis tareas del hogar: hacer la mayor parte de la compra, cocinar, fregar los platos, pasar el aspirador, limpiar el hogar y llevar el presupuesto familiar.

En el primer cuestionario, denominado 'A', para la pregunta sobre satisfacción con la vida y el reparto de tareas en el hogar las opciones aparecen en formato batería, mientras que la presencia de problemas de salud se pregunta en tantas preguntas como problemas. En el cuestionario B se procede de forma inversa, la presencia de problemas de salud aparece en formato batería y las otras dos como preguntas descompuestas (véase el anexo 2).

Con esta forma de proceder, se evita que un cuestionario esté "demasiado cargado" de preguntas de batería y otro sin ninguna. Esto implica que todos los entrevistados van a estar sometidos a alguna pregunta de batería, dos los que responden el cuestionario A y una los que responden el B.

Conviene precisar también que las preguntas no son consecutivas en el cuestionario; se trata de las preguntas 13, 16 y 22 (véase el anexo 2).

Universo y muestra

El universo son las personas de ambos sexos de 18 y más años que residen en hogares de las comunidades autónomas vasca y navarra. Con el fin de obtener un tamaño suficiente para las condiciones planteadas, se seleccionaron dos muestras de 520 personas, muestras estratificadas según zona de residencia y hábitat, y con una selección de las unidades últimas utilizando cuotas de sexo y edad. Las muestras presentan un error muestral del $\pm 4,4$ %, con un nivel de confianza del 95,5 %, $p = 0,5$, y en el supuesto de muestreo aleatorio simple.

Se trata de dos muestras equivalentes (Peterson 2008), considerando las distribuciones en cuanto

a sexo, edad, nivel de estudios, situación de convivencia y tamaño del hogar de la persona entrevistada (véase el anexo 3).

Cada cuestionario fue administrado telefónicamente (383 a números fijos y 137 a móviles) por 22 entrevistadores, con una duración media de 12,8 minutos el cuestionario A y 13,3 el cuestionario B.

El personal que realizó el trabajo de campo recibió una intensa formación que incluyó la aplicación de 16 cuestionarios, 8 de cada versión, entrevistas que fueron grabadas y analizadas por el equipo técnico con el fin de detectar y corregir los errores cometidos. Para reducir al máximo la influencia de los entrevistadores todos aplicaron el mismo número de cuestionarios de cada tipo.

Medidas y análisis de datos

Considerando el planteamiento teórico del *satisficing* o complacencia (Chang y Krosnick 2010; Holbrook, Green y Krosnick 2003; Tourangeau, Rips y Rasinski 2000; Krosnick 1991; 1999; Krosnik et al. 2002), la calidad de respuesta en las preguntas de batería puede definirse operativamente considerando seis criterios: 1) número de ítems no respondidos; 2) elección de respuestas afirmativas en preguntas de escala (aquiescencia); 3) respuesta de las opciones de respuesta extremas; 4) elección de categorías fáciles (posiciones finales y centrales); 5) presencia de correlaciones, y 6) variabilidad (diferenciación) en las respuestas a los ítems que componen la pregunta. Son índices cuyos valores oscilan entre 0 y 1. Se analizará con detalle cada uno, mostrando su cálculo —para el caso de la pregunta 13 en el cuestionario A— en el cuadro 2:

1.º presencia de respuestas no sustantivas, no responder la pregunta u optar por la respuesta 'no sé', definida como *no respuesta parcial* o respuestas "no sustantivas", adoptando la terminología de Peytchev *et al.* (2006). Se trata de un criterio de calidad muy utilizado. De echo, la práctica totalidad de investigación sobre el tema (entre otros, Liu y Cernat 2018; Thorndike *et al.* 2009; Thoepoel, Das y van Soest 2009; Tourangeau, Couper y Conrad 2004; Bell, Mangione y Kahn 2001; Couper, Traugott y Lamias 2001) desvela que colocar los ítems seguidos consigue una mayor rapidez de respuesta, pero esta produce un mayor número de ítems no respondidos. También hay investigaciones (entre otros, Couper, Traugott y Lamias 2001) que detectan resultados opuestos.

El índice se calcula, tal y como se aprecia en el cuadro 2 [fórmula 1], considerando el número de respuestas 'no sabe', que son divididas entre el número de ítems de la pregunta y,

posteriormente, entre el número de casos. Se procede de la misma forma con la falta de respuesta [2]. Tal y como se aprecia en el cuadro 2, en el caso de la batería de la pregunta 13 (tabla 1 en el cuestionario A) se obtienen valores muy bajos, considerando que el índice oscila entre 0 y 1: 0,088 en las respuestas 'no sabe' y 0,18 en la falta de respuesta.

2º Tendencia a responder con opciones favorables o positivas, diferenciando entre *aquiescencia "fuerte"*, para la elección del valor favorable más extremo (el cinco en este caso, véase el anexo 2) y *aquiescencia "suave"*, para elecciones del segundo valor favorable más extremo (el 4). Para el cálculo del primero [fórmula 3 en el cuadro 2], se consideran el

Cuadro 2.
Cálculo de los índices, con aplicación a la pregunta 13, cuestionario A

Ausencia de respuestas	
[1] No sabe =	$\frac{\left(\frac{N.º \text{ de respuestas 'no sabe' (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(322)}{7}\right)}{521} = \frac{46}{521} = 0,088$
[2] No responde =	$\frac{\left(\frac{N.º \text{ de respuestas 'no responde' (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(686)}{7}\right)}{521} = \frac{98}{521} = 0,18$
Número de elecciones de respuestas "positivas" o favorables (aquiescencia)	
[3] Aquiescencia fuerte =	$\frac{\left(\frac{Número \text{ de respuestas 5 (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(1.106)}{7}\right)}{521} = \frac{158}{521} = 0,303$
[4] Aquiescencia suave =	$\frac{\left(\frac{Número \text{ de respuestas 4 (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(1.976)}{7}\right)}{521} = \frac{282,29}{521} = 0,542$
Elección de respuestas extremas	
[5] Respuestas extremas =	$\frac{\left(\frac{Número \text{ de respuestas 1 y 5 (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(1.166)}{7}\right)}{521} = \frac{166,6}{521} = 0,32$
Presencia de "respuestas fáciles" (extremos o el centro de la escala)	
[6] Respuestas fáciles =	$\frac{\left(\frac{N.º \text{ de respuestas 1, 3 y 5 (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(1.290)}{7}\right)}{521} = \frac{184,29}{521} = 0,3537$
Presencia de "respuestas centrales" (centro de la escala)	
[7] Respuestas fáciles =	$\frac{\left(\frac{N.º \text{ de respuestas 3 (prg. 13)}}{Número \text{ de ítems de la pregunta}}\right)}{Número \text{ de casos}} = \frac{\left(\frac{(124)}{7}\right)}{521} = \frac{184,29}{521} = 0,0340$
[8] Elecciones de la misma respuesta en varias preguntas diferentes (ausencia de diferenciación)	
1. Elaboración de tantas variables como respuestas de la escala	
p13 ₁ = Respuestas 'muy insatisfecho' en la pregunta 13	
p13 ₂ = Respuestas 'insatisfecho' en la pregunta 13	
...	
...	
[8] p13 ₅ = Respuestas 'muy satisfecho' en la pregunta 13	
2. Generación de nuevas variables en que se recoge el número de repeticiones	
Para la respuesta 1 "muy satisfecho":	
Cuando p13 ₁ =1 entonces	p13 _{1,1} = 0
Cuando p13 ₁ =2 entonces	p13 _{1,2} = 2/7
Cuando p13 ₁ =3 entonces	p13 _{1,3} = 3/7
Cuando p13 ₁ =4 entonces	p13 _{1,4} = 4/7
Cuando p13 ₁ =5 entonces	p13 _{1,5} = 5/7
Cuando p13 ₁ =6 entonces	p13 _{1,6} = 6/7
Cuando p13 ₁ =7 entonces	p13 _{1,7} = 7/7
Proceder de forma similar en el resto de respuestas 2, 3, 4 y 5 de la pregunta 13	
3. Elaboración del índice	
Diferenciación = [(p13 _{1,1} + p13 _{1,2} + p13 _{1,3} + p13 _{1,4} + p13 _{1,5} + p13 _{1,6} + p13 _{1,7}) + (p13 _{2,1} + p13 _{2,2} + ...) + (p13 _{3,1} + p13 _{3,2} ...) + (p13 _{4,1} + p13 _{4,2} ...) + (p13 _{5,1} + p13 _{5,2} + ...)] / 5	

Fuente: elaboración propia con base en las referencias citadas

- número de respuestas realizadas sobre el valor 5, divididas entre el número de ítems de la pregunta, y el resultado dividido entre el total de casos. En el caso de la aquiescencia “suave”, se considera el valor 4 [4].
- 3° Elección de un mayor número de respuestas extremas, como uno de los rasgos específicos de las encuestas telefónicas (Ye, Fulton y Tourangeau 2011; Nicolaas, Thomson y Lynn 2000; Sykes y Collins 1988; Jordan, Marcus y Reeder 1980; Groves y Kahn 1979). Tal y como se muestra en el cuadro 2 [5], la frecuencia de elección de los valores 1 y 5 es promediada considerando el número de variables que componen la batería. Al igual que los anteriores, se trata de una puntuación que oscila entre 0 y 1.
- 4° *Facilidad de las respuestas*, que añade al indicador anterior el punto medio [fórmula 6 en el cuadro 2], lo que implica la elección de las respuestas más cómodas (extremos) y no definición (punto medio). Se ha elaborado también un índice que recoge el número de elecciones centrales, la no definición [7].
- 5° Presencia de elevadas correlaciones entre los ítems de una batería, debido a que el entrevistado percibe toda la pregunta de una vez. Sudman, Bradburn y Schwarz (1996) consideran que agrupar elementos relacionados en una pregunta aumentará las correlaciones entre ellos, por lo que proponen separar los aspectos considerados en varias preguntas, especialmente cuando se está interesado en el efecto del orden. Asimismo, otras investigaciones realizadas con encuestas autoadministradas a través de Internet localizan mayor correlación entre los ítems cuando se utilizan baterías (Revilla, Toninelli y Ochoa, 2017; Tourangeau, Couper y Conrad 2004; Couper, Traugott y Lamias 2001). Se emplearán, al igual que estos trabajos, las correlaciones de cada ítem con el total (*item-total correlations*), utilizando la Alpha de Crombach. Las preguntas utilizadas presentan cinco y siete ítems, similares a una de las últimas investigaciones publicadas (Silver, Robmann y Gummer 2018), que vuelve a localizar correlaciones más altas en los ítems que forman parte de la batería que en las preguntas individuales.
- 6° Variabilidad de la respuesta del entrevistado a los diferentes *ítems*, denominada ‘ausencia de diferenciación’, entendida como “la utilización de una categoría de respuesta idéntica para todos los ítems de una misma pregunta” (Herzog y Bachman 1981: 551). Detectar la diferenciación es importante porque puede deteriorar tanto la fiabilidad como la validez de

las respuestas de la encuesta, en la medida que su presencia puede “inflar las correlaciones entre los elementos dentro de la batería y, por lo tanto, suprimir las diferencias entre los elementos” (Yan 2008: 521). En este sentido, a juicio de Kim *et al.* (2019), se trata de un aspecto especialmente problemático para las preguntas de batería.

De todos los índices existentes (entre otros, Heerwegh 2009; Narayan y Krosnick 1996; McCarty y Shrum 2000; Alwin y Krosnick, 1991), se ha utilizado la propuesta denominada “método máximo de calificación idéntica” (*Maximum identical rating method*), desarrollada por Holbrook, Green y Krosnick (2003). Para su cálculo¹ se procede, en primer lugar, a elaborar cinco variables (en el caso de una escala de 1 a 5), en que cada una indica el valor de cada posible respuesta [8]. Tomando la nomenclatura del cuadro 2, p13₁ recoge las respuestas ‘muy satisfecho’, p13₂ las respuestas ‘satisfecho’ y, así, hasta los 5 valores de la escala. Posteriormente, se elaboran, para el valor 1 (‘muy satisfecho’) siete variables con un coeficiente que corresponde al número de repeticiones del valor 1. Cuando es elegido una vez (ninguna repetición), la primera variable (p13_{1,1}) es 0; cuando se producen dos elecciones, la segunda variable (p13_{1,2}) es 2/7 (número de repeticiones entre número de ítems); si hay tres la variable es 3/7, y así sucesivamente hasta el 7, mayor número de repeticiones, cuando el valor concedido es 7/7 (véase el cuadro 2). De forma similar, se procede en el resto de respuestas de la escala: p13₂, p13₃, p13₄ y p13₅. Por último, la diferenciación es el promedio, es decir, la suma de todos los valores dividido entre 5.

Valores del índice cercanos a la unidad indican que las personas entrevistadas han expresado las mismas respuestas a todos los ítems de la batería, con independencia de la pregunta planteada. Valores cercanos a 0 están indicando que las personas entrevistadas responden de forma diferente cada ítem, que apenas repiten las respuestas expresadas anteriormente, lo que indica una mayor diferenciación.

Cada una de estas situaciones produce un indicador (de falta de respuesta, aquiescencia, etc.) en cada una de las muestras. Considerando esta situación, con el fin de conocer si existen diferencias significativas en cada medida se utilizó el test de significación de la diferencia de medias, tal y como han procedido investigaciones similares realizadas en otros contextos (Chang y Krosnick 2010; Holbrook *et al.* 2007; Bishop y Smith 2001; Krosnick y Berent 1993). Se utilizará la t de Student cuando la varianza de las variables comparadas

sea igual (homocedasticidad), empleando la *t* de Welch cuando se trate de varianzas desiguales.

Una vez comprobado que existe diferencia entre los indicadores de ambas muestras, es recomendable utilizar alguna medida que proporcione información de la magnitud de esta diferencia, de especial relevancia cuando se trabaja con muestras de elevado tamaño, como es el caso. Es lo que se conoce como ‘tamaño del efecto’. Para conocerlo, se utiliza la *d* de Cohen, que es la diferencia de medias de los dos grupos comparados dividida entre la media de sus desviaciones típicas. Cohen establece tres umbrales para conocer el tamaño del efecto: valores *d* inferiores a 0,20 indican un efecto pequeño, medio si rondan el 0,50, y elevados cuando superan el 0,8 (Pardo y Sanmartín 2010).

RESULTADOS

Índices de calidad

En la tabla 1 se muestran los índices explicados en la sección anterior, aplicados a las tres preguntas analizadas. La primera (pregunta 13) versa sobre el nivel de satisfacción con siete aspectos de la vida del encuestado (salud, vida familiar, relaciones de pareja, amistades, situación laboral, tiempo libre disponible y distribución de las tareas del hogar) y se responde con una escala tipo Likert de cinco opciones, entre ‘muy satisfecho’ y ‘muy insatisfecho’, en que la opción central (‘ni satisfecho ni insatisfecho’) no es ofrecida al entrevistado (véase el anexo 2).

Los índices de calidad de la primera parte de la tabla 1 (pregunta 13) dan cuenta de un mayor número de respuestas ‘no sabe’ en las preguntas de batería. La situación cambia en la falta de respuesta, que es superior en la administración con varias preguntas. El efecto de diseño es pequeño en el primer caso (‘no sabe’), y medio en la falta de respuestas.

Además, la administración en batería proporciona valores superiores en todos los índices, si bien solo cuatro de ellos alcanzan una diferencia significativa. La batería presenta una mayor aquiescencia “fuerte”, mayor elección de respuestas extremas, respuestas fáciles y respuesta central (esta no significativa), así como una mayor correlación entre los ítems. Obsérvese que la diferenciación es elevada, y mayor en las preguntas administradas en batería. Los efectos del diseño son pequeños, rondando el 0,3, como puede observarse en la columna derecha de la tabla 1.

La segunda de las preguntas utilizadas (pregunta 16) versa sobre la frecuencia con la que

han aparecido cinco síntomas relacionados con la salud, y se vuelve a responder con una escala tipo Likert (cinco opciones), si bien en este caso se leen al encuestado todas las opciones de respuesta. Tal y como se aprecia en la parte central de la tabla 1, el número de respuestas “no sustantivas” es muy bajo, presentando la administración en batería un mayor número de no responde (diferencia no significativa). La batería ofrece también una mayor aquiescencia, mayor elección de respuestas extremas, fáciles y respuestas centrales. Esta última es una de las dos que presentan diferencia significativa, con un tamaño del efecto que puede considerarse como pequeño (0,248). Por su parte, la administración en varias preguntas presenta una mayor correlación entre los ítems. También una mayor diferenciación, esta última sin llegar a ser significativa. Obsérvese que esta última presenta magnitudes superiores a las localizadas en la pregunta anterior, consecuencia de leer a la persona entrevistada la respuesta central (‘algunas veces’)². La recomendación es no leer la opción central en este tipo de preguntas.

La tercera pregunta, que corresponde a la número 22 en el cuestionario empleado, plantea el reparto de seis tareas del hogar: comprar, cocinar, fregar, pasar el aspirador, limpiar el hogar y llevar el presupuesto de la casa. La utilización de varias preguntas produce una falta de respuesta ligeramente mayor, si bien el tamaño del efecto es inferior a 0,20 (efecto pequeño). En el resto de indicadores, la pregunta de batería obtiene una peor calidad de respuesta, presentando una mayor aquiescencia, un mayor número de respuestas extremas, de respuestas fáciles, mayor número de respuesta central, mayor correlación entre los ítems y mayor diferenciación, si bien en estas tres últimas la diferencia entre ambas no llega a ser significativa. Es importante señalar los elevados tamaños del efecto de la elección de respuestas extremas, *d* de Cohen de 0,446 rondando el efecto medio, así como la elección de respuestas fáciles (0,368) y la aquiescencia “suave” (0,322).

En síntesis, en la primera pregunta actitudinal se produce una mayor influencia del formato de la pregunta (siete diferencias significativas) que en las de comportamiento, con unas diferencias menores. En la segunda pregunta el formato batería presenta una superior aquiescencia, más respuestas extremas, más respuestas fáciles, centrales, y una mayor diferenciación. Ahora bien, tan solo el rechazo a responder, las respuestas fáciles y la correlación entre los ítems (menor en la batería) presentan diferencia significativa en la aquiescencia “suave”, respuestas extremas, respuestas fáciles y ‘no responde’. La tónica común es una peor respuesta en las preguntas de batería, logrando peores valores en los siete índices considerados.

Tabla 1.
Indicadores de calidad en las preguntas empleadas

¿Nivel de satisfacción con...						
Su salud, vida familiar. Relaciones de pareja, amistades, situación laboral, tiempo libre que tiene y distribución de las tareas del hogar (Preg. 13, 7 ítems)						
5 opciones de respuesta: 1=muy insatisfecho, 5=muy satisfecho						
Tipo de cuestionario						
		Batería	Varias preguntas	Promedio	Sig.	D de Cohen
		Cuestionario A	Cuestionario B			
Falta de respuesta:	No sabe	0,088	0,046	0,067	+ ***	0,292
	No responde	0,188	0,279	0,234	+ ***	0,562
Aquiescencia "fuerte" (5)		0,303	0,274	0,289	+ **	0,287
Aquiescencia "suave" (4)		0,548	0,542	0,545	+	0,291
Elección de respuestas extremas		0,320	0,288	0,304	+ **	0,290
Elección de respuestas fáciles		0,354	0,319	0,336	+ **	0,292
Elección de la respuesta central		0,034	0,031	0,032		0,085
Correlación entre los ítems		0,552	0,517	0,534	**	0,001
Diferenciación		0,730	0,724	0,728		0,176
n		521	519	1.040		
Frecuencia de problemas de salud:						
dificultades en su trabajo por razones de salud, achaques (o dolor en alguna parte de su cuerpo), sentimiento de desdicha o depresión, pérdida de confianza en sí mismo, sentimiento de no poder superar sus problemas (Preg. 16, 5 ítems)						
5 opciones de respuesta: 1=nunca, 5=muy frecuentemente						
Tipo de cuestionario						
		Batería	Varias preguntas	Promedio	Sig.	D de Cohen
		Cuestionario B	Cuestionario A			
Falta de respuesta:	No sabe	0,014	0,015	0,014		0,127
	No responde	0,041	0,025	0,033	+	0,212
Aquiescencia "fuerte" (5)		0,028	0,025	0,027		0,106
Aquiescencia "suave" (4)		0,062	0,057	0,060		0,140
Elección de respuestas extremas		0,641	0,621	0,631		0,322
Elección de respuestas fáciles		0,802	0,776	0,789	**	0,248
Elección de la respuesta central		0,162	0,156	0,159		0,224
Correlación entre los ítems		0,719	0,753	0,736	**	0,001
Diferenciación		0,778	0,766	0,764		0,186
n		519	521	1.040		
Reparto de tareas en el hogar:						
(Preg. 22, 5 ítems)						
5 opciones de respuesta: 1=nunca, 5=muy frecuentemente						
5 opciones de respuesta: 1=entrevistado, 2=pareja, 3=ambos, 4=otros miembros del hogar, 5=otras personas						
Tipo de cuestionario						
		Batería	Varias preguntas	Promedio	Sig.	D de Cohen
Falta de respuesta:	No sabe	0,012	0,014	0,013		0,111
	No responde	0,021	0,037	0,029	+ *	0,178
Aquiescencia "fuerte" (5)		0,041	0,035	0,038	+	0,136
Aquiescencia "suave" (4)		0,151	0,114	0,132	**	0,322
Elección de respuestas extremas		0,589	0,550	0,569	*	0,446
Elección de respuestas fáciles		0,923	0,870	0,896	+ ***	0,368
Elección de la respuesta central		0,334	0,320	0,327		0,361
Correlación entre los ítems		0,810	0,808	0,809		0,000
Diferenciación		0,779	0,767	0,773		0,272
N		521	519	1.040		

Notas:

(+) Varianzas diferentes, utilización de la t de Welch (en el resto t de Student).

(*) Significación al 0,10.

(**) Significación al 0,05.

(***) Significación al 0,01.

Fuente: Elaboración propia.

Distribución de las respuestas

Los índices utilizados desvelan escasas diferencias entre los dos tipos de preguntas, pero pudiera ocurrir que, al tratarse de índices agregados—que consideran la pregunta en conjunto—, dejaran fuera determinados aspectos específicos de las preguntas. Para resolver esta duda, en el gráfico 1 se presenta la distribución de las tres preguntas, diferenciando la administración en batería o en un conjunto de preguntas.

El primer gráfico muestra los valores medios de la pregunta actitudinal, considerando el valor 5 como 'muy satisfecho', el 4 como 'satisfecho', el 3 'ni lo uno ni lo otro', el 2 como 'insatisfecho' y el último (1) como 'muy insatisfecho'. Un análisis visual del gráfico muestra escasas diferencias entre la administración en batería o en varias preguntas, diferencias de una décima que no son suficientemente significativas tras utilizar un contraste de medias.

La diferencia es aún menor en la pregunta sobre la presencia de problemas de salud, en que tres de los cinco ítems presentan la misma media, y la diferencia en los otros dos es, de nuevo, de una décima; por lo que no puede hablarse de una diferencia significativa.

Los ítems de las dos primeras preguntas, al estar medidos con una escala tipo Likert que proporciona una métrica ordinal, permiten trabajar con puntuaciones medias; algo que no es posible con la tercera de las preguntas, en la que cada entrevistado debe identificar la persona que realiza cada una de esas actividades. En este caso, se utiliza la V de Cramer para conocer si el formato del cuestionario influye en la respuesta, y la diferencia de residuos estandarizados corregidos (Haberman 1973) para conocer esa diferencia al nivel de cada celdilla.

Las diferencias de las barras del gráfico no permiten señalar que la distribución de las respuestas sea significativamente diferente cuando se utiliza una pregunta batería o varias preguntas. El análisis celda a celda—siguiendo el procedimiento de Haberman con un valor superior a 1,96—tan solo localiza una mayor asignación por parte del entrevistado en el ítem 'hacer la mayor parte de la compra'. Por su parte, en actividades de limpieza, como fregar los platos o hacer la limpieza del hogar, la utilización de varias preguntas supone un incremento de otros miembros del hogar.

DISCUSIÓN, CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Los resultados obtenidos en la sección anterior ponen de manifiesto que existen diferencias escasas en las respuestas cuando se emplea una pregunta de batería o esta se descompone en tantas preguntas como *ítems* que forman parte de la batería. Las

diferencias son más elevadas en la pregunta sobre actitud que en las dos sobre comportamientos. En este punto, llega el momento de decidir por cuál de ellas optar.

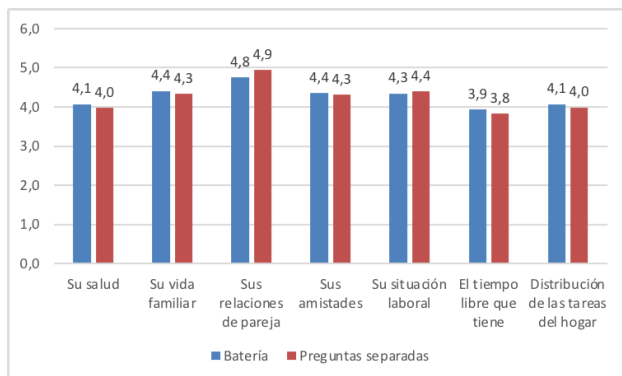
Considerando las ventajas e inconvenientes señaladas en la segunda sección, los datos de la presente investigación coinciden con los que dan cuenta de una administración de cuestionario fácil y rápida (Lozar Manfreda, Vehovar y Koren 2008): el cuestionario con dos baterías (cuestionario A) se responde en 12,8 minutos, y el que tiene una sola pregunta batería en 13,3 minutos (cuestionario B), diferencia significativa. Esta diferencia, lógicamente, se incrementará a medida que se utilicen más baterías. Esto coincide con la opinión de los expertos que las justifican, porque el entrevistado responde más rápido que en otras preguntas (Dillman, Smyth y Christian 2014; Bethlehem y Biffignandi 2011; Couper, Traugott y Lamias 2001).

A estas ventajas, Revilla, Saris y Krosnick (2014) añaden el menor coste de papel y un menor trabajo de los entrevistadores, aspectos no muy relevantes en la medida que—hoy en día—la mayor parte de los cuestionarios son electrónicos (Díaz de Rada 2018); mucho más en el modo telefónico con la implementación de sistemas CATI. Esto supone la eliminación del coste de papel y la reducción de la carga de trabajo de los encuestadores, en la medida que las "páginas" se pasan solas y se cuenta con sistemas de depuración (en tiempo real) y ayudas en la recogida de preguntas abiertas. En relación con el trabajo de los encuestadores, en las sesiones de formación de la presente investigación se les indicó que las baterías debían plantearse haciendo referencia a la pregunta (colocada en la parte superior), esto es, repitiendo las palabras 'satisfacción' o 'insatisfacción' en la pregunta 13 (véase el cuadro 3, en que está todo el texto que debía leerse). Se trataba de lograr la máxima similitud entre los dos tipos de situaciones (batería y varias preguntas), y cumplir así el principio de *estandarización* (Fowler y Mangione 1990), aspecto básico en la técnica de la encuesta. En las sesiones de monitorización de las primeras encuestas, los encuestadores fueron advertidos que debían leer todo el texto de la pregunta mostrado en el cuadro 3.

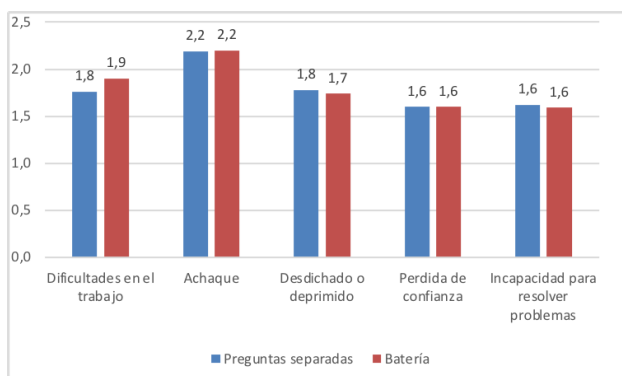
Aunque en las tres preguntas utilizadas las baterías logran mayor diferenciación, las diferencias son escasas y en ninguna de ellas es significativa. Algo similar ocurre con la correlación entre los ítems, superior en las preguntas de batería con dos diferencias significativas. Se ha dado cuenta también de que en las baterías hay una mayor selección de respuestas extremas y fáciles, así como unas respuestas que expresan una mayor tasa de acuerdo con lo planteado. Se confirma, de este modo, las

Gráfico 1.
Distribución de las respuestas

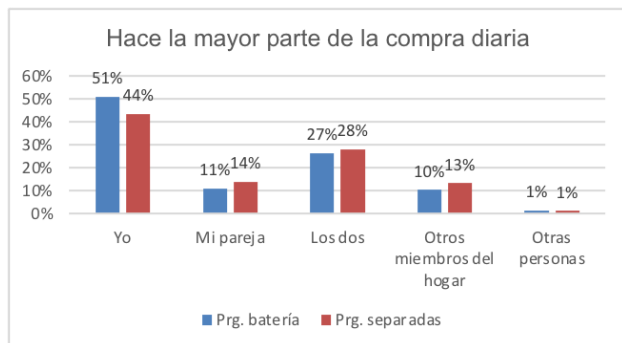
P13. Nivel de satisfacción con 5 aspectos de su vida
(Puntuaciones medias: 1=muy insatisfecho, 5=muy satisfecho)



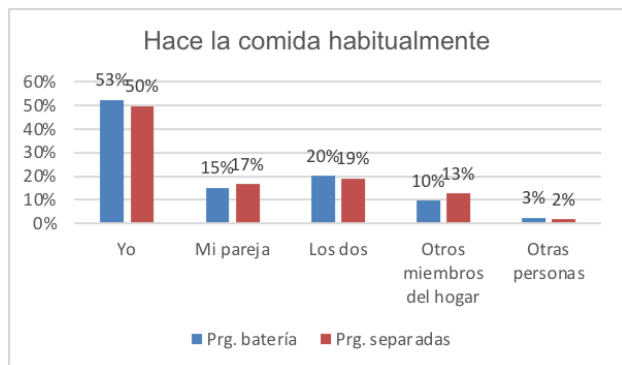
Frecuencia de problemas de salud:
(Puntuaciones medias: 1=nunca, 5=muy frecuentemente)



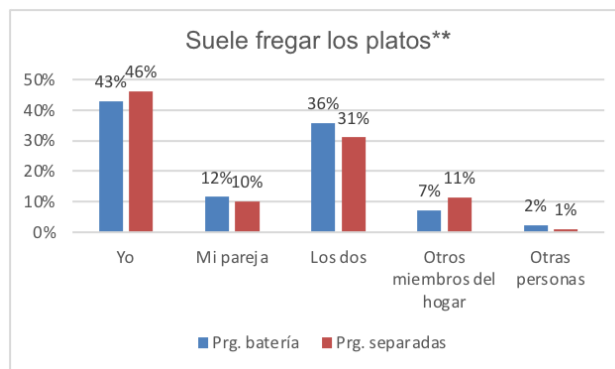
Reparto de tareas en el hogar: (Preg. 22, 5 ítems)
5 opciones de respuesta: 1=entrevistado, 2=pareja, 3=ambos, 4=otros miembros del hogar, 5=otras personas



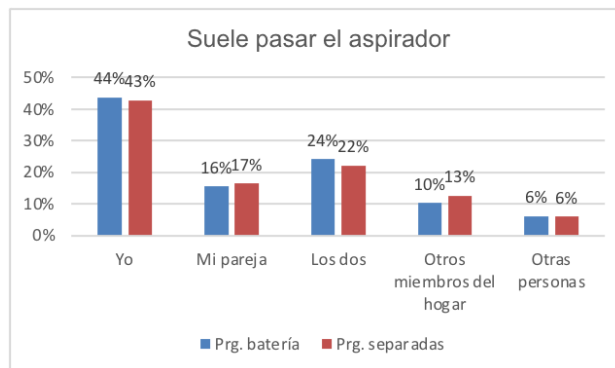
(V de Cramer 0,084, significación 0,117)



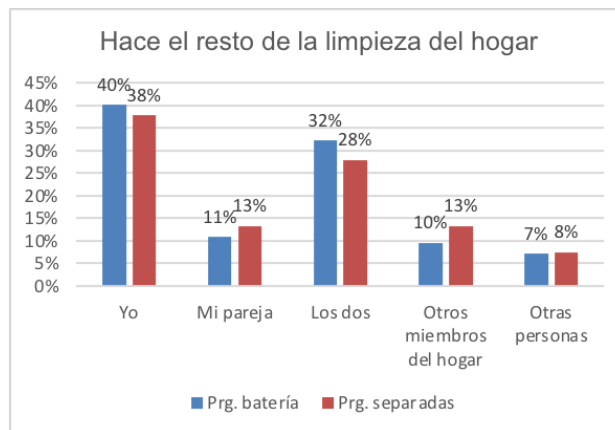
(V de Cramer 0,061, significación 0,575)



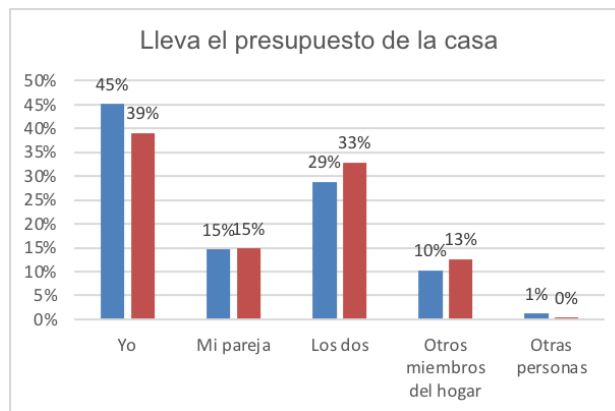
(V de Cramer 0,111, significación 0,056)



(V de Cramer 0,067, significación 0,578)



(V de Cramer 0,084, significación 0,197)



(V de Cramer 0,089, significación 0,227)

Fuente: elaboración propia

Cuadro 3.
Ejemplo de administración de una pregunta de batería

13. ¿En qué medida se siente Ud. satisfecho/a o insatisfecho/a con...

	Muy satisfecho	Satisfecho	Ni satisfecho ni insatisfecho [NO LEER]	Insatisfecho	Muy insatisfecho	No sabe [NO LEER]	No resp. [NO LEER]
A. Su salud <i>Se siente satisfecho o insatisfecho con...</i>	5	4	3	2	1	8	9
B. Su vida familiar <i>Se siente satisfecho o insatisfecho con...</i>	5	4	3	2	1	8	9
C. Sus relaciones de pareja <i>Se siente satisfecho o insatisfecho con...</i>	5	4	3	2	1	8	9
D. Sus amistades <i>Se siente satisfecho o insatisfecho con...</i>	5	4	3	2	1	8	9
E. Su situación laboral <i>Se siente satisfecho o insatisfecho con...</i>	5	4	3	2	1	8	9
F. El tiempo libre que tiene <i>Se siente satisfecho o insatisfecho con...</i>	5	4	3	2	1	8	9
G. La distribución de las tareas en su hogar	5	4	3	2	1	8	9

Fuente: elaboración propia

desventajas del cuadro 1 en lo que respecta a la peor calidad de la respuesta proporcionada a la pregunta de batería.

Al principio de esta sección se planteó que, identificadas las similitudes entre las respuestas a cada tipo de pregunta, es el momento de decidir por cuál de ellas optar. En esta decisión debe considerarse la progresiva aparición de nuevos modos de recogida, generados principalmente por el descenso en la tasa de respuesta (National Research Council 2013). Esta mayor disponibilidad de modos de recogida, unido al hecho de que determinados colectivos colaboran más con unos modos que con otros (Guinaliu Blasco y Díaz de Rada 2021; Dillman, Smyth y Christian 2014), ha llevado a un gran número de expertos (entre otros, De Leeuw 2018; Dillman, Smyth y Christian 2014) a recomendar conjuntamente el empleo de varios modos (modos mixtos) con el fin de lograr una mejor representatividad. El empleo conjunto de varios modos precisa de una adaptación de los cuestionarios a los modos que no cuentan con la ayuda de un encuestador que lo administre, lo que inclina la decisión hacia el uso de varias preguntas en lugar de las baterías. Aunque en el segundo epígrafe se señaló que esta forma de proceder supone más tiempo de administración, también se aludió a la mayor satisfacción de los encuestados (Revilla, Toninelli y Ochoa 2017).

Esta mayor disponibilidad de modos de recogida está generando, a su vez, un desplazamiento de los modos con encuestador (encuestas presenciales y telefónicas) a modos autoadministrados. Utilizando información de Insights+Analytics España, Díaz de Rada (2022) señala que, en el año 2020, el 35 % de las encuestas realizadas por el sector de la investigación de mercados y opinión en España fue realizada *online*, siendo el modo predominante desde el año 2014. En otros contextos, expertos como Kennedy *et*

al (2021) señalan que el 80 % de las encuestas se realizan utilizando modos autoadministrados, dando cuenta de que se trata de una tendencia que se está extendiendo a otros países.

Este predominio del modo autoadministrado, que no se cuenta con un encuestador que ayude al entrevistado, recomienda una eliminación de preguntas complejas como son las baterías. En estas, el encuestado debe cambiar de responder “en vertical” (suponiendo preguntas con opciones de arriba abajo o formato abanico, véase el anexo 2) a responder de forma horizontal y expresar varias respuestas a una misma pregunta. Ahora bien, esta recomendación de preguntas individuales frente a las baterías precisa estudiar la presencia de *primacía* y otros efectos de respuesta en las respuestas individuales, algo que han localizado expertos como Terentev y Maloshonok (2019) y que no se ha abordado en este trabajo. Aparte de esta primera limitación, cabe mencionar una segunda que radica en el hecho de contar únicamente con tres preguntas, una actitudinal y dos de comportamiento, lo que precisa de más trabajos similares con el fin de constatar que los hallazgos aquí localizados son extensibles a otro tipo de temáticas. Una tercera limitación es que en las tres preguntas se utiliza una misma escala de respuestas, escala tipo Likert de cinco opciones etiquetadas, siguiendo así las recomendaciones de Revilla, Saris y Krosnick (2014) cuando indican que se trata de las preguntas con más fiabilidad y validez. La cuarta limitación es que, a pesar de la exhaustiva formación que se le prodigó al personal entrevistador del estudio, no se controlaron aspectos como el tono de voz, el estilo al hablar, el tipo de interacción con el oyente, etc., ni se valoró el cansancio del encuestado con cada pregunta, uno de los aspectos que está reflejado en el cuadro 1.

El hecho de que algunos expertos —entre otros, Liu y Cernat (2018)— señalen que las diferencias entre

las baterías y las preguntas sucesivas se incrementan cuando la escala de respuesta supera las siete opciones justifica la necesidad de seguir trabajando en esta temática. En este sentido, es preciso conocer si escalas más amplias (con más opciones de respuesta) y sin presencia de etiquetas en cada punto presentan los mismos resultados que los localizados aquí. Otro aspecto por considerar es el número de ítems de la pregunta, siete y cinco ítems en el presente artículo, un número muy inferior al utilizado habitualmente en las preguntas de batería (véase el anexo 1).

REFERENCIAS

- Alwin, Duane F. y Jon A. Krosnick. 1985. "The measurement of Values in Surveys: A comparison of ratings and rankings". *Public Opinion Quarterly* 49: 535-552.
- Alwin, Duane F. y Jon A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes". *Sociological Methods and Research* 20: 139-81. <https://doi.org/10.1177/0049124191020001005>.
- Azofra, María José. 1999. Cuestionarios. Madrid: Centro de Investigaciones Sociológicas.
- Bethlehem, Jeike y Silvia Biffignandi. 2011. Handbook of web surveys. Neva York: Wiley.
- Bell, Douglas. Carol Mangione y Charles Kahn. 2001. "Randomized testing of alternative survey formats using anonymous volunteers on the world wide web". *Journal of the American Medical Informatics Association* 8: 616-620. <https://doi.org/10.1136/jamia.2001.0080616>.
- Bishop, George y Andrew Smith. 2001. "Response-Order Effects and the Early Gallup Split-Ballots". *Public Opinion Quarterly* 65: 479-505. <https://doi.org/10.1086/323575>.
- Centro de Investigaciones Sociológicas (CIS) (2022a), Barómetro de enero 2022, estudio nº 3347, https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14604&cuestionario=17590&muestra=25078.
- Centro de Investigaciones Sociológicas (CIS) (2022b), Barómetro de febrero 2022, estudio nº 3351, https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14611&cuestionario=17597&muestra=25091.
- Centro de Investigaciones Sociológicas (CIS) (2022c), Barómetro de marzo 2022, estudio nº 3355, https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14616&cuestionario=17601&muestra=25113.
- Centro de Investigaciones Sociológicas (CIS) (2022d), Barómetro de abril 2022, estudio nº 3359, https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14623&cuestionario=17608&muestra=25132.
- Centro de Investigaciones Sociológicas (CIS) (2022e), Barómetro de mayo 2022, estudio nº 3363, https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14630&cuestionario=17621&muestra=25153.
- Centro de Investigaciones Sociológicas (CIS) (2022f), Barómetro de junio 2022, estudio nº 3366, https://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14634&cuestionario=17635&muestra=25202.
- Chang, Linchat y Jon Krosnick. 2010. "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires, An Experiment". *Public Opinion Quarterly* 74: 154-167. <https://doi.org/10.1093/poq/nfr046>.
- Couper, Mick, Roger Tourangeau., Frederick Conrad y Chan Zang. 2013. "The design of grids in web surveys". *Social Science Computer Review* 31 (3): 322-341. <https://doi.org/10.1177/0894439312469865>.
- Couper, Mick, Michael Traugott y Mark Lamias. 2001. "Web survey design and administration". *Public Opinion Quarterly* 65(2): 230-253.
- de Leeuw, Edith. 1992. Data Quality in Mail, Telephone and Face to Face Surveys. Amsterdam: TT-Publikaties.
- De Leeuw, Edith. y Joop Hox. 2015. "Survey mode and mode effects". Pp. 22-35 en *Improving survey methods. Lessons from recent research*, compilado por Uriel Engel, Benn Jann, Peer Lynn et al. Nueva York and London: Routledge and European Association of Methodology.
- Díaz de Rada, Vidal. 2001. Diseño y elaboración de cuestionarios para la investigación comercial. Madrid: Esic.
- Díaz de Rada, Vidal. 2018. "Encuestas presenciales con cuestionario de papel y ordenador: una comparativa en preguntas de actitudes". *Papers* 103/2: 199-227. <https://doi.org/10.5565/rev/papers.2301>.
- Díaz de Rada, Vidal. 2022. "Strategies to improve response rates to online surveys". *Papers* 107 (4): <https://doi.org/10.5565/rev/papers.3073>.
- Dillman, Don, Jolene D. Smyth y Leath Christian. 2014. Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method (4th edición). Nueva York: Wiley.
- Fowler, Floyd J. y Tom Mangione. 1990. Standarized survey interviewing. Newbury Park, CA: Sage, Applied Social Research Methods Series, vol 18.
- García Ferrando, Manuel y Ramón Llopis. 2015. "La encuesta". Pp. 331-336 en *El Análisis de la Realidad Social*, compilado por Manuel García Ferrando, Francisco Alvira, Louis Enrique Alonso y Modesto Escobar. Madrid: Alianza.
- Graäf, Lorenz. 2002. "Assessing internet questionnaires: The online pretest lab". Pp. 69-79 en *Online social sciences*, compilado por B. Batinic, U.-D. Reips y M. Bosnjak. WA. Seattle: Hogrefe & Huber.
- Greene, Jessica, Howard Speizer y Wyndy Wiitala. 2008. "Telephone and web: Mixed-mode challenge". *Health Services Research* 43: 230-248. <https://doi.org/10.1111/j.1475-6773.2007.00747.x>.
- Groves, Robert y Robert Kahn. 1979. *Surveys by telephone, a national comparison with personal interviews*. San Francisco: Academic Press.
- Guinaliu Blasco, Miguel y Vidal Díaz de Rada. 2021. "Combining sources of information to increase survey response rates". *Spanish Journal of Marketing-ESIC* 25 (1): 29-45. <https://doi.org/10.1108/SJME-04-2020-0060>.
- Haberman, Shelby. 1973. "The Analysis of Residuals in Cross-Classified Tables". *Biometrics* 29: 205-220.
- Heerwegh, Dirk. 2009. "Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects". *International Journal of Public Opinion Research* 21: 111-120. <https://doi.org/10.1093/ijpor/edn054>.
- Herzog, A. Regula y Jerald G. Bachman. 1981. "Effects of questionnaire length on response quality". *Public Opinion Quarterly* 45: 549-559.
- Holbrook, Allyson, Melanie C. Green y Jon Krosnick. 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, Jon Krosnick, David Moore y Rogert Tourangeau. 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>.

- Jordan, Lawrence, Alfred Marcus y Lee Reeder. 1980. "Response styles in telephone and household interviewing, a field experiment". *Public Opinion Quarterly* 44: 102-111. <https://doi.org/10.1086/268585>.
- Kennedy, Courtney, Nicholas Hatley, Arnold Lau, Andrew Mercer, Scott Keeter, Joshua Ferno y Dorene Asare-Marlo. 2021. "Strategies for detecting insincere respondents in online surveys". *Public Opinion Quarterly* 85 (4): 1050-1075. <https://doi.org/10.1093/poq/nfab057>.
- Kim, Yujin, Jennifer Dykema, Paul Moberg, John Stevenson, Penny Black y Paul Moberg. 2019. "Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail-Web Mixed-Mode Surveys". *Social Science Computer Review* 37(2): 214-233. <https://doi.org/10.1177/0894439317752406>.
- 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/10.1002/acp.2350050305>.
- Krosnick, Jon A. 1999. "Survey research". *Annual Review of Psychology* 50: 537-567. <https://doi.org/10.1146/annurev.psych.50.1.537>.
- Krosnick, Jon A. y Duane F. Alwin. 1988. "A test of the form-resistant correlation hypothesis. Ratings, rankings, and the measurement of values". *Public Opinion Quarterly* 52(3): 526-538. <https://doi.org/10.1086/269128>.
- Krosnick, Jon A. y Katthaw Berent. 1993. "Comparison of party identification and policy preferences: the impact of survey question format". *American Journal of Political Science* 37(3): 941-964. <https://doi.org/10.2307/2111580>.
- Krosnick, Jon A., Allyson Holbrook, Matthew Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Koop, Robert Cameron Mitchell, Stanley Presser, Paul A. Ruud, V.Kerry Smith, Wendy R. Moody, Melanie C. Green y Michael Conaway 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>.
- Liu, Mingnan. 2017. "Labelling and Direction of Slider Questions Results from web survey experiments". *International Journal of Market Research* 59(5): 601-624. <https://doi.org/10.2501/IJMR-2017-033>.
- Liu, Mingnan y Alexandru Cernat. 2018. "Item-by-item Versus Matrix Questions: a web survey experiment". *Social Sciences Computer Review* 36(3): 690-706. <https://doi.org/10.1177/08944393166744>.
- Lozar Manfreda, Katja, Vasja Vehovar y Gasper Koren. 2008. "Internet surveys". Pp. 271-283 en *Handbook of public opinion research*, editado por Wolfgang Donsbach y Michael Traugott. Thousand Oaks, CA: Sage.
- McCarty, John A. y Lion Shrum. 2000. "The measurement of personal values in survey research, a test of alternative rating procedures". *Public Opinion Quarterly* 64: 271-298. <https://doi.org/10.1086/317989>.
- Narayan, Sowmya y Jon Krosnick. 1996. "Education Moderates some Response Effects in a Attitude Measurement". *Public Opinion Quarterly* 60: 58-88. <https://doi.org/10.1086/297739>.
- National Research Council. 2013. *Nonresponse in social science surveys: a research agenda*. R. Tourangeau y T. Plewer, editores. Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington DC: The National Academic Press.
- Nicolaas, Gerry, Katarina Thomson y Peter Lynn. 2000. The feasibility of conducting Electoral Surveys in the UK by telephone. Londres: NCSR.
- Olson, Kristen, Jolene Smyth y Beth Cochran. 2018. "Item Location, the Interviewer-Respondent Interaction, and Responses to Battery Questions in Telephone Surveys". *Sociological Methodology* 48 (1): 225-268. <https://doi.org/10.1177/008117501877829>.
- Pardo, Antonio y Rafael San Martín. 2010. *Análisis de datos*. Madrid: Síntesis.
- Peterson, Thomas. 2008. "Split ballots and an experimental approach to public opinion research". Pp. 322-329 en *The SAGE Handbook of Public Opinion Research*, compilado por Wolfgang Donsbach y Michael W. Traugott. Thousand-Oaks, LA: Sage.
- Peytchev, Andy. 2011. "Breakoff and unit nonresponse across web surveys". *Journal of Official Statistics* 27: pp. 33-47.
- Peytchev, Andy, Mick Couper, Sean McCabe y Scott Crawford. 2006. "Web survey design, paging versus scrolling". *Public Opinion Quarterly* 70: 596-607. <https://doi.org/10.1093/poq/nfl028>.
- Revilla, Melanie, Willem Saris y Jon Krosnick. 2014. "Choosing the Number of Categories in Agree-Disagree Scales". *Sociological Methods y Research* 43(1): 73-97. <https://doi.org/10.1177/0049124113509605>.
- Revilla, Melanie, Daniele Toninelli y Carlos Ochoa. 2017. "An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones". *Telematics and Informatics* 34: 30-42. <https://doi.org/10.1016/j.tele.2016.04.002>.
- Revilla, Melanie y Mick Couper. 2018. "Comparing Grids With Vertical and Horizontal Item-by-Item Formats for PCs and Smartphones". *Social Sciences Computer Review*, 36(3): 349-368. <https://doi.org/10.1177/0894439317715>.
- Schuman, Howard. 2008. *Method and Meaning in Polls & Surveys*. Harvard: Harvard University Press.
- Silver, Henning, Joss Robmann y Tobias Gummer. 2018. "When near means related: evidence from three web survey experiments on inter-item correlations in grid questions". *International Journal of Social Research Methodology* 21 (3): 275-288. <https://doi.org/10.1080/13645579.2017.1381478>.
- Sudman, Seymour y Norman Bradburn 2002: *Asking questions*. New York: Wiley.
- Sudman, Seymour, Norbert Bradburn y Norman Schwarz. 1996. *Thinking about answers*. San Francisco: Jossey-Bass.
- Sykes, Wendy y Martin Collins. 1988. "Effects of Mode of Interview, Experiments in the UK". Pp. 30-306 en *Telephone Survey Methodology*, compilado por Robert Groves et al. Nueva York: Wiley.
- Terentev, Evgeniy y Natalia Maloshonok. 2019. "The impact of response options ordering on respondents' answers to rating questions: results of two experiments". *International Journal of Social Research Methodology* 22(2): 179-198. <https://doi.org/10.1080/13645579.2018.1510660>.
- Thorndike, Frances P.; Per Carlbring, Frederick Smyth; Joshua Magee, Linda Gorder-Frederick, Lars, Göran Ost y Lee M. Ritterband 2009. "Web-based measurement: effect of completing single or multiple items per webpage". *Computers in Human Behavior* 25: 393-401. <https://doi.org/10.1016/j.chb.2008.05.006>.
- Toepoel, Vera y Don Dillman. 2011. "How visual design affects the interpretability of survey question". Pp. 165-190 en *Social and behavioral research and the internet*, compilado por Marcel Das, Peter Ester y Lars Kaczmirek. Nueva York.: Routledge <https://doi.org/10.4324/978020384492>.

- Toepoel, Vera, Brenda Vermeeren y Baran Metin. 2019. "Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on Substantive Response, Questionnaire Experience and Response Time". *Bulletin de Méthodologie Sociologique* 142: 57-74. <https://doi.org/10.1177/0759106319834>.
- Toepoel, Vera, Marcel Das, y Arthur van Soest 2009: "Design of web questionnaires: The effects of the number of items per screen". *Field Methods* 21: 200–213. <https://doi.org/10.1177/1525822X08330261>
- Tourangeau, Rogert, Mick Couper y Frederick Conrad. 2004. "Spacing, position, and order: interpretive heuristics for visual features of survey questions". *Public Opinion Quarterly* 68(3): 368-393. <https://doi.org/10.1093/poq/nfh035>.
- Tourangeau, Rogert, Lance Rips y Kennet Rasinski. 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.
- Visauta, Restituto. 1989. *Técnicas de investigación social*. Barcelona: PPU.
- Wojtowicz, Tamara. 2001. "Designing Lengthy Internet Questionnaires: Suggestions and Solutions". Pp. 25-32 en *The Challenge of the Internet; Proceedings of the ASC International Conference on Survey Research Methods*, compilado por Andrew Westlake, Wendy Sykes, Tony Manners y Malcom Rigg. Londres: Association for Survey Computing.
- Yan, Tim 2008. "Nondifferentiation". Pp. 520-521 en *Encyclopedia of survey research methodology*, compilado por Paul Lavrakas. Newbury Park, CA: Sage.
- Ye, Cong, Jenna Fulton y Rogert Tourangeau. 2011. "More positive or More Extreme? A Meta-Analysis of Mode Differences in Response Choice". *Public Opinion Quarterly* 75: 349-365. <https://doi.org/10.1093/poq/nfr009>.

NOTAS

- [1] Agradecemos esta información a la explicación de la profesora Allyson Holbrook, de la Universidad de Chicago.
- [2] En la pregunta 13, la respuesta central es elegida 236 veces, lo cual, considerando las 7290 respuestas (1040*7) supone un 3,24 %. La pregunta 16, por su parte, es elegida 824 veces, lo cual, considerando las 5200 respuestas (1040*5) supone un 16 % de las respuestas.

Vidal Díaz de Rada es catedrático de Sociología de la Universidad Pública de Navarra, director de la Cátedra Opinión Pública e Innovación Social e investigador en I-COMMUNITAS Institute for Advanced Social Research. Sus líneas de trabajo se centran en la *calidad de la encuesta*, mejora del *cuestionario*, *trabajo de campo* y reducción de *no respuesta*, *modos de recogida de información*, y *análisis de datos*. Ha publicado más de dos centenares de trabajos sobre esos temas, así como sobre sociología del consumo.

ANEXO 1: UTILIZACIÓN DE PREGUNTAS DE BATERÍA EN LOS BARÓMETROS DEL CENTRO DE INVESTIGACIONES SOCIOLOGICAS REALIZADOS EN EL PRIMER SEMESTRE DEL AÑO 2022

Barómetros del primer semestre del año 2022:

- Enero (estudio CIS nº 3347), 117 variables de contenido, 49 en formato abanico* y 68 en baterías (**58%**) (Centro de investigaciones sociológicas 2022a).
- Febrero (estudio CIS nº 3351), 90 variables de contenido, 59 en formato abanico y 31 en baterías (**34%**), (Centro de investigaciones sociológicas 2022b).
- Barómetro de marzo (estudio CIS nº 3355), 57 variables de contenido, 24 en formato abanico y 43 en baterías (**75%**) (Centro de investigaciones sociológicas 2022c).
- Abril (estudio CIS 3359), 104 variables de contenido; 24 en formato abanico y 80 en baterías (**77%**) (Centro de investigaciones sociológicas 2022d).
- Mayo (estudio CIS nº 3363), 45 variables de contenido; 26 en formato abanico y 19 en baterías (**42%**) (Centro de investigaciones sociológicas (2022e).
- Junio (estudio CIS nº 3366), 56 variables de contenido; 24 en formato abanico* y 32 en baterías (**57%**) (Centro de investigaciones sociológicas (2022f.)

En definitiva, en cuatro de los seis barómetros analizados hay un mayor predominio de preguntas de batería, superando el 75 % de las preguntas en los barómetros de marzo y abril

* Se adopta este término para referirse a las preguntas de respuesta única que presentan una disposición vertical de las opciones de respuesta. Son preguntas que han recibido diferentes nombres en la literatura, desde 'ratings' (Alwin y Krosnick 1985) o 'forced-choice' (Sudman y Bradburn 1982) en la literatura anglosajona, hasta 'multicategorica' (García Ferrando y Llopis 2015) o, incluso, 'ormato abanico', por la forma en la que se presentan las opciones de respuesta (Visauta 1989). En este trabajo, se opta por este último término.

ANEXO 2: EJEMPLOS DE LAS PREGUNTAS EMPLEADAS

Cuestionario A

13. ¿En qué medida se siente Ud. satisfecho/a o insatisfecho/a con... [LEER CADA UNA]

	Muy satisfecho	Satisfecho	Ni satisfecho ni insatisfecho [NO LEER]	Insatisfecho	Muy insatisfecho	No sabe [NO LEER]	No resp. [NO LEER]
A. Su salud	5	4	3	2	1	8	9
B. Su vida familiar	5	4	3	2	1	8	9
C. Sus relaciones de pareja	5	4	3	2	1	8	9
D. Sus amistades	5	4	3	2	1	8	9
E. Su situación laboral	5	4	3	2	1	8	9
F. El tiempo libre que tiene	5	4	3	2	1	8	9
G. La distribución de las tareas en su hogar	5	4	3	2	1	8	9

Centrados en la salud...

16.A En las últimas cuatro semanas, ¿con que frecuencia ha tenido dificultades en su trabajo (tareas domésticas para los que no trabajan fuera de casa) por razones de salud? [LEER RESPUESTAS]

- Nunca 1
- Rara vez 2
- Algunas veces 3
- Frecuentemente 4
- Muy frecuentemente 5
- No sabe [NO LEER] 8
- No responde [NO LEER] 9

16.B ¿En las últimas cuatro semanas, ¿con qué frecuencia ha tenido un achaque o le ha dolido algo? [LEER RESPUESTAS]

- Nunca 1
- Rara vez 2
- Algunas veces 3
- Frecuentemente 4
- Muy frecuentemente 5
- No sabe [NO LEER] 8
- No responde [NO LEER] 9

16.C En las últimas cuatro semanas, ¿con qué frecuencia se ha sentido desdichado/a o deprimido/a? [LEER RESPUESTAS]

- Nunca 1
- Rara vez 2
- Algunas veces 3
- Frecuentemente 4
- Muy frecuentemente 5
- No sabe [NO LEER] 8
- No responde [NO LEER] 9

16.D En las últimas cuatro semanas, ¿con qué frecuencia ha perdido confianza en sí mismo? [LEER RESPUESTAS]

- Nunca 1
- Rara vez 2
- Algunas veces 3
- Frecuentemente 4
- Muy frecuentemente 5
- No sabe [NO LEER] 8
- No responde [NO LEER] 9

16.E En las últimas cuatro semanas, ¿con qué frecuencia ha sentido que no podía superar sus problemas? [LEER RESPUESTAS]

- Nunca 1
- Rara vez 2
- Algunas veces 3
- Frecuentemente 4
- Muy frecuentemente 5
- No sabe [NO LEER] 8
- No responde [NO LEER] 9

22. ¿Podría decirme quién es la persona que... [LEER CADA UNA]

- A. ...hace la mayor parte de la compra diaria
- B. ...hace la comida habitualmente
- C. ...suele fregar los platos
- D. ...suele pasar el aspirador
- E. ...hace el resto de la limpieza del hogar
- F. ...lleva el presupuesto de la casa

	A	B	C	D	E	F	NS	/NR [NO LEER]
- Yo	1	1	1	1	1	1	8	9
- Mi pareja	2	2	2	2	2	2	8	9
- Los dos	3	3	3	3	3	3	8	9
- Otros miembros del hogar	4	4	4	4	4	4	8	9
- Otras personas	5	5	5	5	5	5	8	9

Cuestionario B

13.A ¿En qué medida se siente Ud. satisfecho/a o insatisfecho/a con su salud? [LEER RESPUESTAS]

- Muy satisfecho5
- Satisfecho4
- Ni satisfecho ni insatisfecho [NO LEER] ...3
- Insatisfecho2
- Muy insatisfecho1
- No sabe [NO LEER]8
- No responde [NO LEER]9

13.B ¿En qué medida se siente Ud. satisfecho/a o insatisfecho/a con su vida familiar? [LEER RESPUESTAS]

- Muy satisfecho5
- Satisfecho4
- Ni satisfecho ni insatisfecho [NO LEER] ...3
- Insatisfecho2
- Muy insatisfecho1
- No sabe [NO LEER]8
- No responde [NO LEER]9

13.C ¿En qué medida... Siguen el resto de aspectos incluidos en la pregunta 13 del cuestionario A.

Centrados en la salud...

16. ¿En las últimas cuatro semanas, ¿con qué frecuencia...? [LEER CADA UNA Y DESPUÉS LAS RESPUESTAS]

	Nunca	Rara- vez	Algunas veces	Fre- cuen- te- mente	Muy fre- cuen- te- mente	No sabe	No resp.
A. Ha tenido dificultades en su trabajo (tareas domésticas para los que no trabajan fuera de casa) por razones de salud	1	2	3	4	5	8	9
B. Ha tenido un achaque o le ha dolido algo	1	2	3	4	5	8	9
C. Se ha sentido desdichado/a o deprimido/a	1	2	3	4	5	8	9
D. Ha perdido confianza en sí mismo	1	2	3	4	5	8	9
E. Ha sentido que no podía superar sus problemas	1	2	3	4	5	8	9

22.A. ¿Podría indicarnos quién hace la mayor parte de la compra diaria? Principalmente Ud., principalmente su pareja, los dos por igual, otros miembros de este hogar u otras personas que no viven en el hogar?

- Yo..... 1
- Mi pareja..... 2
- Los dos por igual 3
- Otros miembros del hogar 4
- Otras personas 5
- No sabe [NO LEER] 8
- No responde [NO LEER] 9

22.B. ¿Podría indicarnos quién hace la comida habitualmente?... Siguen el resto de aspectos incluidos en la pregunta 13 del cuestionario A.

ANEXO 3: COMPOSICIÓN MUESTRAL

Tipo de cuestionario

	A	B	Muestra total
Sexo (V de Cramer 0,019, significación 0,531)			
Hombre	47,8 %	45,7 %	46,6 %
Mujer	52,4 %	54,3 %	53,4 %
Edad (V de Cramer 0,030, significación 0,812)			
De 16 a 30 años	18,8 %	16,8 %	17,8 %
De 31 a 50 años	33,8 %	33,5 %	33,7 %
De 51 a 64 años	24,4 %	25,0 %	24,7 %
65 años y más	23,0 %	24,7 %	23,8 %
Situación de convivencia de la persona entrevistada (V de Cramer 0,062 significación 0,261)			
Soltero/a	23,4 %	22,4 %	22,9 %
Casado/a y viviendo en pareja	67,0 %	66,4 %	66,7 %
Separados/as y divorciados/as	4,6 %	3,5 %	4,0 %
Viudos/as	5,0 %	7,7 %	6,4 %
Nivel de estudios (V de Cramer 0,52 significación 0,579)			
Primarios incompletos	10,9 %	13,9 %	12,4 %
Primarios completos	20,0 %	21,2 %	20,6 %
Secundarios	33,6 %	31,8 %	32,7 %
Universitarios medios y universitarios sin finalizar	14,4 %	12,9 %	13,7 %
Licenciado/graduado y más	21,1 %	20,2 %	20,7 %
Tamaño del hogar de la persona entrevistada (V de Cramer 0,058 significación 0,472)			
Unipersonal	10,9 %	12,0 %	11,5 %
Dos miembros	34,7 %	33,7 %	34,2 %
Tres miembros	20,2 %	23,0 %	21,6 %
Cuatro miembros	26,3 %	22,2 %	24,3 %
Entre 5 y 8 miembros	7,9 %	9,1 %	8,5 %
N.º de casos	521	519	1040