

## La senda tortuosa de la «calidad» de la encuesta<sup>1</sup>

M.<sup>a</sup> Ángeles Cea D'Ancona

Universidad Complutense de Madrid

macda@cps.ucm.es

### RESUMEN

En el actual contexto social de búsqueda y exigencia de *calidad*, la encuesta no puede ser ajena. La credibilidad y significatividad de los datos que aporta la encuesta, en el análisis de la realidad social, están supe-  
ditadas al rigor que se haya puesto en el diseño, ejecución e interpretación de la información recogida. El presente artículo sintetiza, de una manera com-  
prehensiva, plural y diacrónica, los aspectos a consi-  
derar en la evaluación de la *calidad* de la encuesta, subrayando los *errores* que pueden confluír en su rea-  
lización. Se parte de una visión conjunta y entrelaza-  
da de los *errores* de encuesta, de la necesidad de su conocimiento, en aras de su predicción y control. La concreción de los *límites* de la encuesta, junto con las actuaciones preventivas y paliativas en su resolución, contribuyen al avance en su *calidad*. Razón por la cual se recogen los pasos más consensuados en la firme andadura por la senda de la *calidad* de la encuesta.

**Palabras clave:** Métodos Cuantitativos, Encuestas, Evaluación de Calidad, Análisis de Errores.

### ABSTRACT

In the present social context of search and demand for *quality*, the survey cannot be absent. The credibility and significance of the data that the survey provides in the analysis of social reality depends of the level of rigor placed on the design, execution and interpretation of the information collected. This article synthesizes, in a comprehensive, plural and diachronic way, the aspects to be considered in the evaluation of survey *quality*, underlining the *errors* that can come together in its execution. We start from a joint and intertwined vision of survey *errors*, as well as from the need to know them, for the sake of their prediction and control. The concreteness of the *limits* of the survey, together with the preventive and mitigating actions in their resolution, contribute to the advance towards its *quality*. This is the reason why we show the steps most scholars agree with in relation to the firm tread along the path of survey *quality*.

**Keywords:** Quantitative Methods, Surveys, Quality Evaluation, Errors Analysis.

<sup>1</sup> Este artículo arranca de una primera aproximación, presentada como ponencia marco en el VIII Congreso Español de Sociología (celebrado en Alicante, del 23 al 25 de septiembre de 2004), con el título *El reto de la calidad de la investigación mediante encuesta*.



La búsqueda de la *calidad* es una exigencia cada vez más demandada por la sociedad contemporánea. En cualquier producto que se oferta en el mercado se busca el «sello» que acredite que cumple requisitos mínimos de «calidad». Su estampación da al usuario más confianza sobre el bien que adquiere. La demasía de mercancías y la competitividad entre ellas, para alcanzar cotas de mercado, hacen que se eleven las exigencias en las mismas y sus costes se relacionen directamente con la *calidad* del producto. Aunque no siempre los *costes* más elevados son garantía de *calidad*. Ésta no sólo depende de factores económico-materiales, también cuenta el factor humano que interviene en su producción y comercialización.

La encuesta, como bien de uso, no se libra de esta creciente demanda de *calidad* que también se exige de la información que aporta al conocimiento de la realidad social<sup>2</sup>. Las asociaciones ESOMAR (Asociación Europea de Estudios de Mercado y Opinión) o AEDEMO (Asociación Española de Estudios de Mercado, Marketing y Opinión) persiguen ser garantes de «calidad» en los estudios de mercado, como es AENOR (Asociación Española de Normalización y Certificación) en los productos comerciales en general. En 1948, ESOMAR edita un primer *Código de Práctica Profesional*, un *corpus* de normas deontológicas y prácticas deseables que perseguía el «control» de los estudios de mercado para transmitir *confianza* al usuario. En 1977, ESOMAR y CCI (la Cámara de Comercio Internacional) elaboran un código conjunto, que sería posteriormente revisado en 1986 y 1994<sup>3</sup>.

Pero ¿qué se entiende por *calidad* en la investigación mediante encuesta? ¿Cuándo puede decirse que una encuesta cumple criterios de calidad? A estos y otros interrogantes se trata de dar respuesta en las páginas que siguen. Especial mención se hará a los *errores* que pueden surgir en cada fase en el desarrollo de la encuesta, sus repercusiones en sus resultados y qué hacer en su resolución en aras de mejorar la *calidad* de la encuesta.

<sup>2</sup> En España, el Instituto Nacional de Estadística (INE), por ejemplo, aplica, desde 1971, un programa de evaluación de calidad en la *Encuesta de Población Activa* (EPA). La finalidad es disponer de información sobre los errores de muestreo y los ajenos al muestreo que permita mejorar la encuesta, además de proporcionar al usuario información sobre la calidad de los datos que ofrece.

<sup>3</sup> AEDEMO y ANEIMO (Asociación Nacional de Empresas de Investigación de Mercados y de la Opinión Pública) son dos asociaciones españolas que aplican el *Código Internacional ESOMAR-CCI*, de Prácticas de Marketing e Investigación Social. En Estados Unidos, CASRO (Council of American Survey Research Organization) edita en 1998 su propia guía para la investigación mediante encuesta (<http://www.casro.org/guidelines.cfm>). Abarca las distintas facetas en el proceso de encuesta: el diseño de la muestra, de la entrevista, la recogida de datos (formación del entrevistador, supervisión), su procesamiento (edición, codificación, entrada de datos, depuración y tabulación) y el informe de la encuesta.

## 1. LA ELABORACIÓN DE UN MARCO CONCEPTUAL EN LA EVALUACIÓN DE LA CALIDAD DE LA ENCUESTA

En 1915, Gillin ya alertaba de los peligros de la falta de controles de *calidad* en la encuesta. Un año después, en 1916, Hobson inicia una serie de estudios experimentales dirigidos a comprobar los efectos del cambio (en los cuestionarios) de la formulación de la pregunta. Culminan en los experimentos dirigidos por Cantril (1944) desde The Office of Public Opinion Research, creado por él mismo en la Universidad de Princeton<sup>4</sup>, quien también analiza los sesgos debidos al entrevistador; y en la más reciente publicación de Schuman y Presser (1981), de gran repercusión en la literatura posterior sobre encuesta.

A partir de investigación propia, complementada por la llevada a cabo por otros investigadores, un total de 130 experimentos en 34 encuestas nacionales y regionales, entre 1971 y 1980, Schuman y Presser analizan los efectos debidos a la forma de la pregunta, su redacción y contexto en los resultados de encuesta. Constatan que pequeños cambios en la formulación de las preguntas pueden provocar grandes diferencias en las respuestas. Con anterioridad, Sudman y Bradburn (1974) habían demostrado que los efectos de la formulación de la pregunta afectan tanto a las preguntas de opinión como a las de hecho. A esta conclusión llegaron tras la realización de un meta-análisis de 900 experimentos *split-ballot*<sup>5</sup> que muestra que la mayoría de los *errores de medición* se deben al cuestionario: su extensión, el orden de las preguntas, su formato abierto o cerrado, el empleo de palabras ambiguas o de difícil comprensión, la saliencia de la pregunta para el encuestado y el empleo de ayudas al recuerdo.

Pero no es el *error de medición* el más medido en las encuestas, sino el *error muestral*. Este último, el único que se recoge en la *ficha técnica* (siguiendo la recomendación de ESOMAR-ICC), que se demanda como garante de «calidad». Una genérica y escueta descripción de quién hace la encuesta, en qué fecha, dónde (el ámbito geográfico de la investigación) y de quién (la población de estudio). Se indica el tamaño global de la muestra, el *error muestral* para unos márgenes de confianza y de heterogeneidad determinados (generalmente estimado a la baja, aplicando las fórmulas correspondientes al muestreo aleatorio simple, que beneficia a los diseños muestrales polietápicos por conglomerados) y el procedimiento seguido en la selección de la muestra. En la ficha ni siquiera se recoge

<sup>4</sup> Hadley Cantril también diseñó varios de los aproximadamente 400 experimentos sobre los efectos de la redacción de las preguntas en los cuestionarios elaborados por Gallup entre 1936 y 1949.

<sup>5</sup> El procedimiento experimental más aplicado en la comprobación de diferencias en las respuestas ante dos alternativas de una misma pregunta. Para ello se divide la muestra al azar en dos mitades, o más, dependiendo del número de formulaciones alternativas que se haga de la misma pregunta. Garantizada la homogeneidad de los grupos, por la distribución aleatoria de sus integrantes, las diferencias en las respuestas se interpretan como debidas a la formulación de la pregunta.

información relativa a los otros *errores de no observación* (errores de *no cobertura* y de *no respuesta*), que afectan igualmente a las posibilidades de inferencia y de generalización de las estimaciones muestrales. Y la dada sobre el *error muestral* en general se estima insuficiente.

Pese al protagonismo del *error muestral*, el análisis de la precisión de las estimaciones muestrales también exige calibrar los *errores de no cobertura* (a quiénes de la población de estudio se ha excluido en la selección de la muestra) y de *no respuesta* (la composición de la muestra final y su correspondencia con la población de interés). Todos ellos afectan a la representatividad de la muestra y a las posibilidades de inferencia de los datos de encuesta, siendo el *error muestral* insuficiente en su cuantificación.

Asimismo, tampoco ha de menospreciarse la contribución del *entrevistador* a la *calidad* de la encuesta (*cara a cara* y *telefónica*). Ya en 1929, Rice hablaba de «sesgos de contagio» debidos a la intervención de los entrevistadores en el proceso de recogida de información. En 1942, Katz demuestra que la falta de homogeneidad de estatus social entre el entrevistador y el entrevistado contribuye a la obtención de respuestas sesgadas: los entrevistadores de «cuello blanco» (de clase social media-alta) recaban opiniones, actitudes y conductas más conservadoras que los de «cuello azul» (de clase baja-trabajadora).

En 1944, Deming incluye los errores debidos al *entrevistador* en su listado de trece factores que deterioran la calidad de la encuesta (junto con los errores de muestreo, de no respuesta, los debidos al procedimiento de recogida de datos, al diseño del cuestionario, la codificación de las respuestas, el procesamiento de la información y su interpretación). Dos años después, Mahalanobis publica uno de los primeros estudios experimentales sobre la contribución del entrevistador a los errores de encuesta, mediante la asignación aleatoria de entrevistadores en una encuesta agrícola y el uso de «muestras interpenetradas»<sup>6</sup>, para la estimación de la varianza de los entrevistadores en las distintas variables incluidas en el cuestionario. Destaca la necesidad de incluir los «errores humanos» (aquellos debidos a la actuación de entrevistadores, de codificadores y de supervisores), junto a los errores de muestreo, en la estimación del *error de encuesta*.

En 1954, Hyman y colaboradores analizan cómo las expectativas del entrevistador afectan a los datos de encuesta, en un estudio comparativo en las ciudades de Memphis y Nueva

---

<sup>6</sup> Mahalanobis defiende el uso de la *replicación*, el «interpenetrar muestras» para medir la precisión de las estimaciones derivadas de una encuesta. La *replicación*, el repetir la operación de muestreo, el extraer una serie de submuestras comparables, ya fue inicialmente propuesta por A. N. Kiaer (el entonces director de la Agencia Estadística Noruega de Oslo) en 1897, el primer defensor del uso de *muestras representativas* en la investigación mediante encuesta. Sus propuestas fueron posteriormente desarrolladas por J. Neyman (1934), quien demuestra que el *error de muestreo* puede medirse calculando la varianza del estimador.

York. Las *expectativas* de qué respuestas van a obtener influyen en su actuación durante la entrevista y la obtención final de las respuestas que esperaban conseguir. Además, constataron que características sociodemográficas (etnia, religión, género) del entrevistador y del entrevistado contribuyen a la captación de datos de encuesta «sesgados»<sup>7</sup>.

Tres años después, Kahn y Cannell resumen en tres las fuentes originarias del *error* debido al *entrevistador*: sus características sociodemográficas (género, edad, nivel educativo, estatus socioeconómico, etnia, religión), factores psicológicos (las expectativas o las percepciones sociales<sup>8</sup>) y una actuación inadecuada en la realización de la entrevista<sup>9</sup>.

La investigación empírica más reciente corrobora estas fuentes originarias de error debidas al entrevistador, aunque con algunas matizaciones. Las *características sociodemográficas* más visibles (género, edad, etnia) afectan, pero cuando estén relacionadas con el tema de la encuesta. Aunque sus efectos se anulan cuando se controla por la *experiencia* del entrevistador<sup>10</sup>. Ésta es la variable que parece más determinante en la consecución de un elevado porcentaje de respuesta y de mayor calidad (Groves y Fultz, 1985; Groves y Lyberg, 1988; Feldman, Hyman y Hart, 2002; Biemer y Lyberg, 2003). Si bien se matiza que existe una relación curvilínea entre actuación del entrevistador y la experiencia (Biemer y Lyberg, 2003), y que los entrevistadores «muy experimentados» tienden a ser descuidados, incluso más que los «noveles» (Fowler, 1993). Además, afecta la forma de mirar del entrevistador, de vestirse, de hablar y de adaptar las instrucciones a cada encuestado.

Respecto a los factores psicológicos, se destacan dos: la *confianza* que el entrevistador tenga en sí mismo y sus *expectativas* sobre los encuestados. Los entrevistadores que tie-

<sup>7</sup> Obtienen que los entrevistados negros dan respuestas más patriotas, en preguntas referidas a política exterior, cuando el entrevistador es blanco, o se autclasifican en un estatus social más elevado si responden a entrevistadores negros. Los efectos debidos al entrevistador fueron mayores en Memphis que en Nueva York.

<sup>8</sup> En el transcurso de la entrevista, el entrevistado puede *percibir* que el entrevistador espera de él unas respuestas determinadas y, en consecuencia, ajustar sus respuestas a aquellas que reflejan las *expectativas* de lo que considera respuestas «correctas». Coincide con lo que Orne (1969) llama «características de la demanda» como determinante fundamental de la conducta de los entrevistadores y los entrevistados.

<sup>9</sup> En España, un estudio de Camacho y colaboradores (2000) detecta una serie de irregularidades muy graves en la actuación de los entrevistadores dignas de considerar en la planificación de una encuesta.

<sup>10</sup> En una encuesta telefónica sobre acoso sexual se obtuvo que los encuestados varones eran dos veces más probables a informar haber acosado sexualmente en el trabajo cuando eran entrevistados por un hombre que por una mujer (Lavrakas, 1998). Igualmente, en una encuesta que medía actitudes de rol de género, Kane y Macaulay (1993) encontraron efectos en la respuesta debidos al sexo del entrevistador. Cuando éste era mujer, tanto los entrevistados varones como las mujeres daban respuestas más igualitarias que cuando el entrevistador era varón. En el meta-análisis de Sudman y Bradburn (1974) los mayores errores se detectaron en entrevistadores de menos de 25 años, en su mayoría estudiantes universitarios. La explicación no está en la edad, sino en la carencia de la necesaria *experiencia* y *formación* como entrevistadores que garantice una actuación correcta.

nen más *confianza* en su capacidad para conseguir cooperación y completar con éxito la entrevista acaban consiguiendo una mayor cooperación por parte de los encuestados y una mayor tasa de respuesta (Groves y Lyberg, 1988; Frey y Oishi, 1995; Groves y Couper, 1998). Las *expectativas* en relación a los resultados de la encuesta pueden, igualmente, afectar a la actuación del entrevistador e incidir en la captación de respuestas coincidentes con sus expectativas: la llamada *profecía autocumplida*<sup>11</sup>. A estas variables se suman las comúnmente relacionadas con una actuación adecuada: su *preparación* o formación para el trabajo a realizar; su *interés* y *motivación* (en la que incide negativamente la baja remuneración y la precariedad en el trabajo); el *tiempo* que disponga para su realización; la *cantidad de trabajo* que tenga que realizar y su *compatibilidad* con otras ocupaciones (contrato a jornada parcial).

Pero los «errores humanos» no se reducen al entrevistador. También están los *errores* debidos al *encuestado*. Principalmente, los ocasionados por su *voluntad* de dar respuestas veraces (en especial en preguntas que puedan afectar a su autoestima o imagen pública) y/o su *incapacidad* para dar información correcta. Ya sea debido a una mala redacción de la pregunta (que favorece su errónea comprensión), ya a su desconocimiento del tema que se le pregunta, a su precipitación en la respuesta o a fallos en su memoria. Todo lo cual se traduce en *errores de medición*, en la obtención de respuestas que se desvían de los valores verdaderos en la población. En su valoración, lógicamente, habrá asimismo que considerar los posibles efectos de la mediación del entrevistador (cuando esté presente). A éstos se sumarán los propiciados por la presencia de otras personas durante la cumplimentación del cuestionario, además de los acontecimientos que ocurrieron antes y durante su realización.

También están los *errores* que resultan de actuaciones incorrectas por parte de los *procesadores* y *analistas de datos*, que tratan de paliarse con la creciente automatización del procedimiento de encuesta (sistemas CADAC<sup>12</sup> o los modos automatizados de grabación mediante escáner, fibra óptica, grabación oral). Estos errores no son siempre reconocidos en la comprobación de la *calidad* de la encuesta, pese a que pueden traducirse en *errores sistemáticos*<sup>13</sup> (los más nocivos), porque esta fase suele recaer en pocas

<sup>11</sup> Bradburn y Sudman (1979) lo comprueban en una encuesta sobre consumo de drogas, alcohol y conducta sexual a 1.172 adultos norteamericanos. Antes de la realización del trabajo de campo preguntaron a cada uno de los 59 entrevistadores por el grado de dificultad que esperaban que tendría la encuesta y la proporción de encuestados que responderían a la pregunta. Tras la realización de la encuesta encontraron una pequeña relación, aunque consistente, entre *expectativas* previas y datos obtenidos. Aquellos entrevistadores que preveían dificultad en la administración de los cuestionarios y que no lograrían mucha información fueron precisamente quienes menos información consiguieron.

<sup>12</sup> Referenciados genéricamente bajo las siglas CADAC (*Computer-Assisted Data Collection*), CASIC (*Computer-Assisted Survey Information Collection*) o CAI (*Computer-Assisted Interviewing*).

<sup>13</sup> La distinción entre *errores sistemáticos* (aquellos que no suman cero cuando se promedian las observaciones muestrales, por lo que habrá una subestimación o sobreestimación de los valores «reales» en la población) y *errores variables* (sí su-

manos. Además, no todos los *errores de grabación* y de *codificación* son fácilmente perceptibles<sup>14</sup>.

En suma, en la realización de una encuesta convergen distintas fases a cargo de personas concretas. Éstas pueden adoptar, intencionadamente o no, actuaciones «erróneas», que no sólo afectan a su hacer, sino también a fases posteriores en el curso de la investigación, deteriorando su *calidad*. Bien por la presencia de los llamados *errores de no observación* (de no cobertura, de muestreo y de no respuesta), porque no se ha recogido información de toda la variedad de unidades que componen la población de estudio; bien por *errores de medición* u observación. Estos últimos relativos a la adecuación de los datos que la encuesta aporta, a su correspondencia con la «realidad». Las desviaciones de las respuestas pueden deberse al *método de encuesta* (cómo se aplique el cuestionario), al *diseño del cuestionario* (las preguntas, su disposición, presentación), a la intervención del *entrevistador* (características personales y su actuación durante la entrevista), a la actitud del *encuestado* (estado de ánimo, cansancio, interés, la imagen que quiera proyectar de sí mismo, sus conocimientos del tema, la memoria) y al *tratamiento de la información* (codificación, grabación, análisis estadísticos).

Todos ellos se consideran integrantes básicos en cualquier evaluación de la *calidad* de la encuesta. Así lo hace Groves (1989), en una publicación de gran impacto en la literatura de encuesta, o, más recientemente, Biemer y Lyberg (2003), quienes también sintetizan las distintas fuentes de error, insistiendo en su visión conjunta, y la existencia de una relación inversa entre la *calidad* y el *coste* de la encuesta.

Sánchez Carrión (1996) amplía la perspectiva global del análisis de la *calidad* de la encuesta. Ha de incluir un «triple ojo escrutador»: el *técnico* (que analiza los posibles errores cometidos en su implementación), el *metodológico* (que mira la pertinencia de su utilización) y el *epistemológico* (que observa la encuesta después de haber contestado una pregunta clave: ¿para qué sirve la investigación, para *explicar* la realidad social, para *comprender* los fenómenos sociales o para *transformar* el objeto de estudio?). Los tres niveles se hallan relacionados en orden jerárquico, de manera que la respuesta que se dé al *epistemológico* influirá en el *metodológico* y éste, a su vez, en el *técnico*. Se precisa conocer las posibilidades y los límites de la encuesta; su adecuación como recurso metodológico al objeto de estudio.

---

man cero cuando se promedian) se debe a Hansen, Hurwitz y Madow (1953). Estos autores también proponen el concepto de *error total de encuesta*, desde cuya óptica ha de analizarse el *error de encuesta*.

<sup>14</sup> Biemer y Lyberg (2003) estiman que los *errores de codificación* pueden llegar a niveles del 20% cuando el personal de codificación no está debidamente formado. Campanelli y colaboradores (1997) relacionan los *errores de codificación* con la conjunción de factores adversos: el tipo de pregunta, la naturaleza de la respuesta, la longitud y la adecuación del marco de codificación, la formación y la supervisión de los codificadores.



Tampoco ha de descuidarse el *contexto social* en el que se aplica la encuesta<sup>15</sup>. Álvarez (2004) diferencia tres planos. Uno, el conocimiento de la *sociedad* donde se aplica. Se destaca el efecto negativo de la «saturación» de encuestas y de la demanda de información que se solicita a los encuestados. Dos, las relaciones entre el ámbito de demanda y el científico. En el *ámbito de demanda* están los agentes de demanda directos (quienes determinan los objetivos de la encuesta y su uso, aportando los recursos económicos para su realización) y los *usuarios* (los que utilizan los datos de encuesta). Estos últimos requieren información sobre cómo se ha realizado la encuesta para poder valorar la calidad de los datos. Como antes se ha apuntado, una *ficha técnica* con el mayor detalle posible. Y, por último, la interacción social que se establece entre el entrevistador, el investigador y el entrevistado, que no siempre están debidamente interrelacionados<sup>16</sup>. Es queja habitual la falta de interrelación entre quienes diseñan la encuesta y aquellos que la ejecutan.

En esta reflexión sobre los determinantes de la encuesta hay que sumar el punto de vista de los tres componentes básicos en cualquier proceso investigador: los *objetivos* perseguidos, los *recursos* o medios materiales, económicos y humanos de que se disponga para hacer factible el estudio, y el *tiempo* concedido para su realización. Obviamente, a mayores recursos y plazo temporal para su ejecución, se espera una mayor *calidad* de la investigación. Pero necesariamente no es así. También influye la *formación* o cualificación de todos los intervinientes en el proceso investigador, tanto en la fase de diseño como en la recogida de datos, en su análisis e interpretación. Del equipo investigador se espera que *diseñe* una buena encuesta, que satisfaga los objetivos de la investigación y que se adecue a los recursos y al plazo temporal dados; que sea capaz de prever los distintos *errores* que pueden surgir en la encuesta y adopte medidas preventivas para evitarlos. De los ejecutores, que se ajusten a las directrices marcadas en el diseño de la encuesta y, en caso de desviarse, que informen al equipo investigador para su consideración en el análisis y correcta interpretación de los datos que la encuesta aporta. El reconocimiento de «errores» y la delimitación precisa de los «límites» de la encuesta son cruciales en la interpretación del alcance de los resultados y contribuyen a su *credibilidad*. Ésta suele vincularse a la obtención de datos *válidos* y *fiables*.

---

<sup>15</sup> La importancia del *contexto social* ya fue resaltada por Cialdini (1984), Groves, Cialdini y Couper (1992) o por Groves y Couper (1998), en su explicación de la participación en la encuesta. También es reconocida en la *Encuesta Social Europea* como requisito de *calidad*. La consideración del *contexto* en el que se realiza el estudio, junto con la equivalencia en la redacción de las preguntas y la uniformidad de las opciones de respuesta, contribuyen a la calidad en estudios internacionales (Riba y Cuxart, 2004).

<sup>16</sup> Fernández Esquinas (2003) defiende igualmente la necesidad de aunar los componentes *científicos* y *técnicos* en el proceso de investigación, con las condiciones sociales objetivas en las que ésta se desarrolla. La necesidad de considerar el *contexto social* en el que se produce la investigación.

## 2. ACTUACIONES EN LA MEJORA DE LA ENCUESTA

La satisfacción de criterios de *validez* y de *fiabilidad* compartidos en la metodología cuantitativa tradicionalmente se vincula a la consecución de una encuesta de «calidad». Ha de cubrir los *objetivos* del estudio, pero también los criterios de *validez* inicialmente propuestos por Campbell y Stanley (1970) y posteriormente modificados por Cook y Campbell (1977) y Reichardt y Cook (1979). Es decir, *validez interna* (el «control» de explicaciones alternativas en las relaciones causa-efecto analizadas), *externa* (la generalización de los datos de encuesta, que depende de la presencia de los *errores de no observación*), de *conclusión estadística* (circunscrita al análisis estadístico de los datos) y de *constructo* (o adecuación de la medición de los conceptos principales del estudio, lo que dependerá de la existencia de *errores de medición* y de *especificación*). Su cumplimiento es garantía de *fiabilidad*, pero no a la inversa.

La *fiabilidad* se relaciona con la «estabilidad» y la *consistencia* de los datos. Por lo que la forma más habitual de comprobar la *fiabilidad* es mediante la aplicación reiterada de un mismo cuestionario, en distintos momentos, para observar si se alcanzan los mismos resultados<sup>17</sup>. Si bien su no consecución no siempre significa que el cuestionario sea «infiable». Es el analista quien deberá precisar las causas.

Como con la *validez*, hay que considerar las diversas fuentes de *error* que pueden determinar la *fiabilidad* de la encuesta. Afectan los *errores de medición*, pero también los de *no observación*. La proporción y las características diferenciales de los «excluidos» o «no observados» respecto a los que sí han participado en la encuesta. La obtención de un coeficiente de *fiabilidad* elevado se relaciona con la observación de *muestras* amplias y heterogéneas respecto al objeto de estudio; la *operacionalización múltiple*<sup>18</sup> (el empleo de múltiples indicadores); la eliminación de aquellos *indicadores* no relevantes, que repercute en incrementos en el coeficiente *alfa*<sup>19</sup>; el aumento de la *varianza* de las variables (incrementando las opcio-

<sup>17</sup> Uno de los procedimientos que aplica el INE en su evaluación de la calidad de la *Encuesta de Población Activa* es, precisamente, repetir la entrevista poco tiempo después de la entrevista original, a una parte de la muestra, siguiendo el modelo de Hansen, Hurwitz y Bershady y que aplica la Oficina del Censo de Estados Unidos. Asimismo, en la *Encuesta Social Europea* se ha optado por incluir algunas de las preguntas del cuestionario principal, con otros formatos, en un cuestionario complementario que se administra al final de la entrevista. Con ello se quiere evaluar la *validez* y la *fiabilidad* de los indicadores. También el *análisis factorial* es de gran utilidad en la cuantificación de la *consistencia interna*, y el *análisis discriminante* permite comprobar la efectividad de cada ítem, por separado, en la diferenciación de los sujetos, de acuerdo con el concepto que se mida. Como muestra Cea D'Ancona (2002) en la medición del concepto de racismo.

<sup>18</sup> Inicialmente propuesta por Campbell y Fiske (1959) como estrategia para aumentar la *validez de constructo*.

<sup>19</sup> Propuesto por Cronbach en 1951, aunque Kuder y Richard (1937) fueron los primeros en recomendar una medida de equivalencia que utilizase todas las covarianzas entre los ítems. La *no relevancia* suele medirse con un valor bajo de *comunalidad* (o valor  $R^2$ : el coeficiente de correlación múltiple cuadrado obtenido de regresionar la variable con el resto de las incluidas en el análisis).

nes de respuesta de las preguntas); y la *estandarización* del proceso de recogida de información, para que se reduzca la varianza debida al entrevistador. También, el emplear un mayor número de entrevistadores contribuye a que su efecto en la *calidad* de la encuesta repercuta en un menor número de cuestionarios (traduciéndose en *error aleatorio* o *variable* y no en *sistemático*). O, como dicen Henerson, Morris y Fitz-Gibbon (1987: 149):

«La mejor forma de demostrar que tu trabajo ha sido mínimamente contaminado por inconsistencias de “instrumentos humanos” es utilizar más de una persona para hacer al menos una muestra de tus entrevistas u observaciones. Si personas diferentes informan bastante la misma cosa, eso es evidencia de consistencia».

El cuadro 1 resume las actuaciones más consensuadas en la reducción de los *errores de no observación*. Como en éstos afecta no sólo la tasa de población finalmente «no observada», sino, muy especialmente, sus características diferenciales con aquella sí observada, las actuaciones se dirigen a reducir ambos componentes de *error*.

Como la *no cobertura* está propiciada por la utilización de listados con erróneas inclusiones u omisiones de unidades de la población (ya por su no actualización, por duplicidades

CUADRO 1

## Remedios ante los errores de no observación

| Errores de encuesta | Remedios  |
|---------------------|---|
| De no cobertura     | <ul style="list-style-type: none"> <li>— Evaluar la «calidad» del marco muestral</li> <li>— Restringir la población de estudio a la población <i>marco</i></li> <li>— Utilizar conjuntamente varios marcos muestrales</li> <li>— Identificar las «omisiones» y utilizarlas como variables de <i>cuota</i></li> <li>— Aplicar otros métodos de encuesta</li> <li>— Ponderar la muestra</li> </ul>  |
| De muestreo         | <ul style="list-style-type: none"> <li>— Incrementar el tamaño de la muestra</li> <li>— Selección plenamente aleatoria que garantice la heterogeneidad de la muestra</li> </ul>   |
| De no respuesta     | <ul style="list-style-type: none"> <li>— Adecuar el trabajo de campo a la población. Si fuese necesario, ampliar su duración</li> <li>— Prenotificación de la encuesta</li> <li>— Seguimiento de la muestra</li> <li>— Gratificar la participación</li> <li>— Mejorar la presentación de la encuesta</li> <li>— Emplear otro método de encuesta en los no respondientes</li> <li>— Convertir las «negativas» en aceptación</li> <li>— Incrementar el tamaño de la muestra</li> <li>— Sustituir a los no respondientes por unidades equiparables</li> <li>— Ajustes estadísticos postencuesta mediante la ponderación</li> </ul> |

o por su no correspondencia con el objeto de estudio), la primera actuación para evitar *errores de no cobertura* supone la evaluación de la «calidad» del marco muestral<sup>20</sup> y su adecuación a la población de estudio. En la medida en que los excluidos o «no cubiertos» en el marco presenten características distintas a los incluidos, las estimaciones de la encuesta estarán sesgadas. En caso contrario, el *sesgo* apenas será perceptible. Ésta ha sido la razón principal del mayor *error de no cobertura* tradicionalmente atribuido a la encuesta telefónica y, en la actualidad, a las encuestas *DBM (Disk By Mail)*, *EMS (Electronic Mail Survey)* y las *encuestas Web*<sup>21</sup>. En caso de detectar carencias en la cobertura del *marco muestral*, las actuaciones más comunes son:

- Restringir la definición de la población de estudio a la incluida en el marco muestral porque el error desaparece cuando ambas poblaciones coinciden.
- Utilizar «conjuntamente» varios marcos muestrales, cuando se prevea que uno solo no cubre la población de estudio.
- Identificar las «omisiones» del marco y utilizarlas como variables de *cuota* en el diseño de la muestra (para garantizar su presencia).

En referencia a la encuesta telefónica, Keeter (1995) propone el uso de «hogares transitorios» (aquellos que han experimentado un cambio en su situación de usuario de telefonía fija: acaban de incorporarse al directorio telefónico o han dejado de estar registrados) para conocer las características de los hogares «sin teléfono».

- Aplicar otros métodos de encuesta para conocer la cobertura (telefónica o de los modos informatizados de encuesta) y las características diferenciales de la población cubierta y la no cubierta en las variables concretas que se analizan. Este remedio supondría, obviamente, incrementar los costes de la investigación. Aunque éstos podrían, en

<sup>20</sup> La «calidad» de los *marcos muestrales* resulta muy dispar cuando se cotejan en estudios que abarcan distintos países. Sea, por ejemplo, en el *Eurobarómetro 53: Attitudes towards minority groups in the European Union* (SORA, 2001), en especial en lo referido a las direcciones. Las que resultan erróneas varían desde el 0% en Alemania, hasta el 61% en España o el 56% en Finlandia. Porcentajes demasiado elevados que indudablemente sesgan los resultados de la investigación. En la *Encuesta Social Europea* (uno de nuestros referentes porque persigue la *calidad* de la encuesta) se siguen detectando diferencias en la calidad de los marcos muestrales, aunque en menor magnitud. En España se seleccionan direcciones del listado de secciones censales y aquellas que resultaron inválidas se redujeron al 13% (Riba y Cuxart, 2004).

<sup>21</sup> El no tener teléfono está diferencialmente relacionado con el siguiente perfil sociodemográfico: un nivel educativo y de renta bajo, residir en hábitats rurales, edad más temprana del cabeza de familia y pertenecer a minorías étnicas (Couper y Hansen, 2002). En España, Díaz de Rada (2001) lo relaciona con el nivel de ingresos, el número de personas en el hogar y el tamaño del municipio de residencia. En la más reciente *Encuesta AIMC*, a usuarios de Internet, de febrero-marzo de 2004, se obtuvo el siguiente perfil de internauta: varón (57,6%), entre 25-34 años (31,5%), de clase media-media (42,5%). Ello limita su aplicación en encuestas a la población general. No cuando se dirija a un grupo de población concreto (profesores universitarios, profesionales), en su mayoría usuarios de la red, al menos en sus lugares de trabajo.

parte, amortizarse si la información referida a la población no cubierta pudiese utilizarse en encuestas posteriores «próximas» en el tiempo.

- Ajustes estadísticos postencuesta mediante la *ponderación* para reducir los efectos del *error de no cobertura* en las estimaciones muestrales, al igual que se hace en los otros *errores de no observación*. Ello exige disponer de información «actualizada» de la población no cubierta. Su éxito dependerá de lo relacionadas que estén las variables utilizadas en la ponderación con el problema de investigación, y de que realmente diferencien a la población cubierta de la no cubierta.

El *error de muestreo*, el tradicionalmente más medido en la encuesta, requiere para su reducción, no para su eliminación<sup>22</sup>, incrementar el tamaño de la muestra y aplicar diseños muestrales plenamente aleatorios que garanticen la heterogeneidad de su selección<sup>23</sup>. Pero aunque se haya diseñado una muestra «perfecta», que cumpla los criterios exigidos de tamaño y de aleatoriedad en la elección de sus integrantes, si finalmente no se consigue que éstos respondan el cuestionario, teniéndose que recurrir a «otros» para obtener información, todos los recursos a ella destinados se habrán malogrado. Los datos que se extraigan de la encuesta, aunque satisfagan requisitos de significatividad estadística (al haberse mantenido el tamaño de la muestra), no serán «representativos» de la población de interés, si no se mantiene la composición de la muestra original e intervienen otros factores ajenos a la aleatoriedad.

Uno de los retos más difíciles a los que se enfrenta la encuesta en la actualidad es alcanzar la plena participación de los elegidos para ser encuestados. A ello contribuye una serie de factores que tienen que ver con:

- a) *Contexto social*. Groves, Cialdini y Couper (1992), junto con Groves y Couper (1998), destacan la contribución de factores relacionados con la *responsabilidad social* sentida por la persona y la *legitimidad*<sup>24</sup> que ésta atribuya a la encuesta. Con posterioridad, Curtin, Presser y Singer (2000) insisten en el sentimiento de *obligación cívica* en la explicación de

<sup>22</sup> Porque siempre habrá *error muestral* al extraerse información sólo de una parte de la población y ésta únicamente representa una de todas las posibles muestras que pueden seleccionarse de la población.

<sup>23</sup> En los años noventa han ido apareciendo programas especializados en el diseño de muestras complejas para la estimación de errores de muestreo por conglomerados (Wesvar) o procedimientos de *remuestreo* (Bootstrapp), que permiten solventar el problema del error muestral mediante la *replicación* artificial de un número elevado de muestras a partir de la original (Fernández Esquinas, 2003: 65). Pero no siempre logran calibrar todos los errores muestrales.

<sup>24</sup> A la *legitimidad* contribuye la temática de la encuesta y el prestigio del centro responsable. Por esta razón, es práctica usual, en la presentación de la encuesta, mencionar el propósito del estudio, quién lo lleva a cabo e, incluso, el patrocinador, si su conocimiento se piensa que propicia la participación en la encuesta.

la participación en la encuesta. También afectan la *saturación* o exceso de encuestas realizadas, la *imagen* que se tenga de ellas y los *cambios sociales* que dificultan la participación. A destacar la creciente incorporación de la mujer al mundo laboral, el estar menos tiempo en casa, el mayor *individualismo* de las sociedades modernas, el *rechazo* a hablar con extraños y el *temor* al uso que se hará de la información dada.

b) *Método de encuesta*. La *no respuesta*, tanto *total* como de *ítem* (o pregunta), suele ser menor en la encuesta *cara a cara* y mayor en la encuesta por correo<sup>25</sup>. El contacto visual con el encuestado contribuye a su *motivación* (para participar en la encuesta) y a una mejor explicación del cuestionario, que ayuda en la reducción de la *no respuesta* de *ítem*. Además, se detecta una *no respuesta* distinta en cada método de encuesta.

En la *encuesta telefónica* la *no respuesta* se debe más a «rechazos» que a «no contactos», por el mayor seguimiento de las unidades de la muestra (la *rellamada*). En la encuesta *cara a cara* el mayor coste de las *revisitas* incide en que éstas se reduzcan a una, dos o a lo más tres (como es deseable), resultando en una mayor *no respuesta* por «no contacto». Y, una vez contactada la persona, el decir «no» a la cara de un entrevistador es más difícil que colgar el teléfono o no rellenar el cuestionario. En la *encuesta por correo* es más complicado conocer el motivo de la *no respuesta*: no contacto (dirección errónea, ausencia del domicilio) o rechazo.

En la *encuesta telefónica* las negativas suelen darse en el primer minuto de conversación. Después de que los entrevistadores se han presentado, pero antes de que se explique el propósito de la llamada<sup>26</sup>. La encuesta *cara a cara* es, en cambio, la menos vulnerable a la *no respuesta* parcial (el interrumpir la entrevista mientras se está realizando).

En relación con la encuesta también afecta el *tema* que se trate. No tanto si es un tema sensible o delicado (Singer, 1978; Bradburn y colaboradores, 1978), sino el *interés* que suscite en el encuestado. Si éste cree que tiene información relevante que pueda aportar, en consonancia con la *teoría de la saliencia* desarrollada por Groves, Singer y Corning (2000).

<sup>25</sup> En el meta-análisis realizado por Hox y De Leeuw (1994, 2002), de 45 estudios que aplicaron distintas modalidades de encuesta, se obtuvo, como promedio, que las encuestas *cara a cara* conseguían un porcentaje de respuesta más elevado (70%). Le seguía la encuesta *telefónica* (67%) y, a mayor distancia, la encuesta por *correo* (61%). Este gradiente coincide con la revisión hecha años antes por Goyder (1987), quien, tras comparar 385 encuestas por correo, 112 *cara a cara* y 53 telefónicas, obtuvo las siguientes tasas de respuesta: 67,3% en *cara a cara*, 67,3% en *telefónica* y 58,4% por *correo*.

<sup>26</sup> En el estudio de Oksenberg y Cannell (1988), con entrevistadores telefónicos del Survey Research Center de Michigan, el 40% de las negativas se produjeron durante las primeras frases en la presentación de la encuesta; un 50%, después de las primeras frases; el 10% restante, una vez comenzada la entrevista. En la revisión de cuatro encuestas telefónicas hecha por Collins *et al.* (1988), en el Reino Unido, un tercio de las negativas ocurrieron incluso antes de que el entrevistador pudiese establecer qué individuo, de la vivienda, debía ser entrevistado.

c) *Características de la persona a encuestar.* El perfil del no respondiente suele identificarse más con personas mayores, de un nivel educativo medio-bajo, que viven solas, en áreas urbanas y son varones (Traugott, 1987; Davis y Smith, 1992; Singleton y Straits, 2002)<sup>27</sup>. Aunque hay un cierto disenso a este respecto.

d) *Características del entrevistador.* Su sexo, edad, etnia, estatus social, e incluso su afinidad política, pueden afectar a la participación en la encuesta, cuando interactúan con *prejuicios* y preferencias de los encuestados y tienen, además, relación con el tema de la encuesta. A éstos se suman sus *rasgos psicológicos* (expectativas de éxito, percepciones sociales, confianza en sí mismo, motivación), las *exigencias laborales* (cantidad de trabajo, remuneración), su *preparación* para la encuesta concreta a realizar, su *manera de hablar*<sup>28</sup> o de expresarse verbalmente (más en la *encuesta telefónica*), de vestir, de mirar (en la *encuesta cara a cara*), y su *experiencia* (que contribuye favorablemente a la reducción de la *no respuesta*), como ya se indicó. Su *paciencia* y la *actitud*, en general, durante la entrevista. En mayor medida, cuando se encuesta a personas con dificultades de lenguaje o cognitivas que impiden un desarrollo fluido de la entrevista. Y, por último, *cómo se enfrenten a la no respuesta*, su perseverancia y capacidad disuasoria para convertir la *no respuesta* por *rechazo* en respuesta (como muestran Bishop, Tuchfarber y Oldendick, 1986).

Debido a que en el *error de no respuesta* inciden tanto la tasa de *no respuesta* como las características diferenciales entre quienes responden y quienes no, su *reducción* también estará marcada por ambos componentes. Sin duda, la consecución del primero, una tasa de *no respuesta* prácticamente nula, favorece lo segundo (por ser mínima la presencia de no respondientes en la *muestra*). Por esta razón, los principales esfuerzos se dirigen a conseguir que los sujetos aleatoriamente elegidos para ser encuestados acaben participando en la *encuesta*. Diez son las actuaciones principales, como se sintetiza en el cuadro 1:

— *Adecuación del trabajo de campo a las peculiaridades de la población a encuestar*<sup>29</sup> y, si fuese necesario, *ampliación del período de tiempo para el trabajo de campo*. Posibili-

<sup>27</sup> La incidencia diferencial del tamaño de hábitat, en contra de los núcleos más urbanos, también se detecta en los barómetros del CIS (Núñez, 2004), ya por una mayor desconfianza hacia «extraños», por más barreras de acceso a los edificios, o porque los estilos de vida se traducen en una mayor dificultad para establecer contacto. Lo mismo se observa en la *Encuesta Social Europea* (Riba y Cuxart, 2004): una tasa de respuesta del 49% en el ámbito urbano (capitales de provincia) y del 57% en el ámbito rural (resto de municipios). El mayor rechazo también se da en personas de 40 a 59 años (34%) y en varones (29%), más que en mujeres (25%). La participación más alta, en los jóvenes de 15 a 19 años (90%) y en los mayores de 60 años (75%).

<sup>28</sup> Oksenberg y Cannell (1988) destacan la conveniencia de hablar en voz alta, con decisión (sin titubeos) y con variabilidad en el tono de voz. Contribuye a que se perciba a los entrevistadores como más competentes y creíbles. De lo que resulta una mayor colaboración por parte de los encuestados.

<sup>29</sup> En el estudio de Trujillo y Serrano (2003) se muestra la distinta probabilidad de contactar por teléfono a varones y a mujeres en días y franjas horarias diferentes. Las mujeres, mejor de 12 a 21 horas, al igual que los varones mayores de 60 años.

ta el seguimiento de los «no localizados» en distintos días y horas de la semana, antes de la aplicación del remedio radical de la *sustitución* para no reducir el tamaño de la *muestra*.

- *Prenotificación de la encuesta*. El preaviso (por teléfono o carta) puede contribuir más a la reducción de la *no respuesta* por «rechazo» que por «no contacto». Aunque el preaviso telefónico ayude más a identificar problemas en la ejecución del trabajo de campo, la prenotificación por carta es más habitual y suele asociarse con una mayor tasa de respuesta<sup>30</sup>. Cuando la *prenotificación* se realice *por carta*, se aconseja que incluya el *membrete oficial* de la agencia patrocinadora porque ayuda en la *legitimidad* de la encuesta. Más en el caso de que tenga autoridad y sea de reconocido prestigio (Brunner y Carroll, 1969; Groves y Lyberg, 1988)<sup>31</sup>.
- *Seguimiento de la muestra*. Ante la «no contactación», lo más conveniente es insistir en la llamada, en distintos días y horas del día (en la *encuesta telefónica* o *cara a cara*) o enviar, de nuevo, el cuestionario (en la *encuesta por correo*)<sup>32</sup>. Son tales los beneficios en la reducción de la tasa de respuesta por «no contacto», que se aconseja incluir «la planificación de rellamadas al calcular los costes y el tamaño de la muestra, al diseñar la muestra y los procedimientos de campo» (Kish, 1965/1995: 551). La facilidad de su aplicación difiere en los distintos métodos de encuesta. De mayor aplicación, por su menor coste, en la *encuesta telefónica*, y más restrictiva en la *encuesta cara a cara*<sup>33</sup>.

Los varones, en general, a las 21 horas. Localizar a las mujeres los sábados era más difícil que los días laborales, tanto por la mañana (excepto a primera hora) como por la tarde. El domingo fue el día más favorable para contactar a varones jóvenes. Sobre todo, a última hora de la mañana y de la tarde. Aunque fue el día de la semana en que menos varones respondieron a la encuesta. El efecto de la *estacionalidad* es demostrado por Vigderhous (1981) y Losch *et al.* (2002).

<sup>30</sup> Cifrada, en la encuesta por correo, en el 47,4% (Fox, Crask y Kim, 1988) o 47,9% (Faria, Dickinson y Filipic, 1990). En la encuesta telefónica, con aumentos en la tasa de respuesta de 5 a 13 puntos porcentuales (Traugott, Groves y Lepkowski, 1987). Pero no tanto si la *prenotificación* de la *encuesta telefónica* se hace por *teléfono*. Groves y Magilavy (1981) probaron un contacto telefónico antes de una encuesta aleatoria RDD (*Random-Digit-Dialing*). Las diferencias en los porcentajes de respuesta apenas fueron perceptibles: 81% en aquellos en los que sí hubo un contacto previo y del 80% en los que no hubo.

<sup>31</sup> En el estudio de Brunner y Carroll (1969), cuando la *prenotificación* era remitida por una universidad, la tasa de respuesta aumentaba del 52,3 al 72,5%. En cambio, cuando la enviaba un organismo privado, el incremento se reducía sensiblemente del 47,7 al 53,9%.

<sup>32</sup> Heberlain y Baumgartner (1978) muestran que el *número de contactos*, junto con la *saliencia* o relevancia del tema, explican el 51% de las variaciones en la *tasa de respuesta*.

<sup>33</sup> Lavrakas (1993) documenta la necesidad de incrementar las *rellamadas* en la *encuesta telefónica*. Con una única llamada registra una tasa de respuesta del 27%; con cinco llamadas, del 77%, y con siete, del 86%. En España, en el estudio de Trujillo y Serrano (2003) se logró contactar al 90% de las mujeres con más de 45 años con seis llamadas telefónicas. Los hombres menores de 45 años, los más difíciles de contactar, precisaron nueve llamadas para alcanzar ese porcentaje de éxito. En la *encuesta cara a cara*, Kish (1965/1995) aconseja repetir la visita seis o más veces, hasta alcanzar una *tasa de respuesta* del 85%. Como en la *encuesta telefónica*, en la *encuesta cara a cara* también conviene llevar un registro de las *revisitas* efectuada para agilizar su realización.



En la encuesta telefónica también se aconseja mantener el teléfono sonando el máximo tiempo posible y acortar la presentación de la encuesta (Dillman, Gallegos y Frey, 1976; Couper y Hansen, 2002), para que el entrevistado no pierda interés y rehúse participar.

En la *encuesta por correo*, la efectividad de los *reenvíos* de cuestionarios ha sido ampliamente documentada. Ya en 1961, Scott comprueba que los envíos de seguimiento eran lo más efectivo para incrementar la *tasa de respuesta* y, en 1978, Dillman los incluye en su propuesta para aumentar la respuesta en la *encuesta por correo: el Método de Diseño Total (TDM: Total Design Method)*. Mayor efectividad se detecta si el reenvío se hace por correo certificado. Suele duplicar la tasa de respuesta respecto al correo convencional<sup>34</sup>. Además, permite identificar a los «no respondientes», de interés en la valoración de la *no respuesta*. Otra cuestión es la respuesta que pueda dar el encuestado con tanta presión para que responda. Puede reducirse el *error de no respuesta* y, por el contrario, incrementarse el de *medición*. También, el alargamiento en exceso del período de trabajo de campo puede afectar negativamente a la *comparabilidad* de los datos.

- *Gratificar la participación*. Partiendo de la *teoría del intercambio*, Dillman (1978, 1991) propone «gratificar», aunque sea simbólicamente, la participación en la encuesta como estrategia para incentivar la *tasa de respuesta*, al menos en la *encuesta por correo*. Con anterioridad, Gouldner (1960) había señalado la norma social de la «reciprocidad» como base para el uso de *incentivos* en la encuesta. Es preferible un regalo «simbólico», de escaso valor (de 5 euros, *por ejemplo*), adjunto al cuestionario, que una *gratificación* cuantiosa (un coche, 6.000 euros, un viaje para dos personas) «prometida» tras la cumplimentación del cuestionario. Los *incentivos monetarios* son más convincentes que los no monetarios (Goodstadt *et al.*, 1977; Hansen, 1980); más los *prepagados* que la promesa de recompensa condicionada a la cumplimentación del cuestionario (Wotruba, 1966; Armstrong, 1975; Gelb, 1975; Goodstadt *et al.*, 1977; Berk *et al.*, 1987)<sup>35</sup>. En su contra está, lógicamente, el incremento en los costes de la encuesta y, como apunta

<sup>34</sup> Aumento del 43% en la tasa de respuesta frente al 20% en correo ordinario (House, Gerber y McMichael, 1977). En los más reacios, el seguimiento telefónico puede ser necesario (Brennan y Hoek, 1992).

<sup>35</sup> Con una *gratificación* de sólo 5 dólares dada por adelantado, James y Bolstein (1990) obtienen una tasa de respuesta del 71,3%, mientras que con la promesa de 50 dólares sólo lograron un 56,7%. Además, consiguieron una información de mayor *calidad* (escasa *no respuesta de ítem* y preguntas *abiertas* más completas). Con posterioridad, Warrimer y colaboradores (1996), igualmente con un prepago de 5 dólares en una *encuesta por correo*, lograron una tasa de respuesta del 78%. Mangione (1998), en una encuesta sobre consumo de alcohol, llegó a una tasa de respuesta del 82% con la misma gratificación. Con un prepago de sólo 1 dólar, pero diez años antes, Hubbard y Little (1988) habían conseguido una tasa de respuesta del 68%, mientras que con la promesa de gratificación de 200 dólares la tasa de respuesta se redujo al 51,8%. En una serie de experimentos realizados durante dos años, en una *encuesta telefónica* mensual (*Survey of Consumer Attitudes*), aplicando el procedimiento *RDD (Random-Digit-Dialing)*, Singer, Van Hoewyk y Maher (2000) consiguieron incrementar la tasa de respuesta en al menos 10 puntos porcentuales con *incentivos prepagados* incluidos en la *carta de preaviso*. No con incentivos prometidos ni con cartas de notificación sin gratificación incluida.

Sánchez Carrión (2000), la posibilidad de modificar la respuesta, al intentar los encuestados «agradar» con sus respuestas. Pero esto último no ha sido corroborado por la investigación empírica (Hubbard y Little, 1988; James y Bolstein, 1990).

- *Mejorar la presentación de la encuesta.* Se insiste en la necesidad de procurar garantías de *anonimato* y de *confidencialidad*. En especial, cuando la encuesta aborde cuestiones «delicadas» que puedan suscitar suspicacias por parte del encuestado<sup>36</sup>. También es importante destacar la *relevancia* de la información que el encuestado pueda aportar y los *beneficios* de su participación.
- *Emplear otro método de encuesta en los no respondientes.* Para los «no contactados» en la *encuesta cara a cara* puede concertarse, por *teléfono*, una «cita». Si el entrevistado lo prefiere, incluso puede hacerse la entrevista por teléfono. Igualmente, con quienes no devuelven el cuestionario cumplimentado por *correo*, si se dispone de un número de teléfono. O enviar *por correo* el cuestionario a los «no contactados» vía *cara a cara* o *telefónica* (si se tiene su dirección). Y, en el caso de la *encuesta cara a cara*, si hay alguien en el domicilio, se le puede entregar el cuestionario para que se lo dé a la persona a la que ha de encuestarse y ésta lo rellene cuando pueda, en un plazo de tiempo marcado. El cuestionario puede recogerse en una fecha y hora concretas o, si se prefiere, se les entrega un sobre contra reembolso para que lo envíen por *correo*. De esta forma se reducirían los costes económicos y temporales en el seguimiento de los no respondientes. Aunque puede haber problemas de *comparabilidad* en las respuestas, por el modo diferente de cumplimentarse el cuestionario.
- *Tratar de convertir las «negativas» en aceptación.* De acuerdo con Van Kammen y Stouthamer-Loeber (1998), casi la mitad de aquellos que se niegan a ser entrevistados después consienten en serlo. Por lo que se aconseja reasignar los casos «reacios» a entrevistadores sénior, enviarles una carta de «persuasión» o cambiar el método de encuesta (Groves y Couper, 1998). Pero el forzar a los reacios puede introducir *errores de medición*, si éstos no dan respuestas precisas (por no querer pensar, por querer ocultar la respuesta u otra causa)<sup>37</sup>.

<sup>36</sup> Singer, Von Thurn y Miller (1995) demuestran que la garantía de *confidencialidad* es lo que más favorece la respuesta en *encuestas «delicadas»*. Si bien advierten que su uso «en exceso» en las demás encuestas puede tener el efecto contrario: aumentar, en el encuestado, la sensación de tema muy delicado. De nuevo, la recomendación de que en la medida está el acierto. Con anterioridad, Singer (1993/2002) había concluido que el *anonimato* y la *confidencialidad* son claves para aumentar la *participación* en la encuesta y la *calidad* de la respuesta. No obstante, la variable crucial es la *confianza* en la integridad de la agencia que recoge los datos, no la naturaleza de las *garantías* dadas a los respondientes.

<sup>37</sup> Davis y Smith (1992) observan que aquellos que al principio se negaron a realizar la entrevista, pero que después fueron convencidos, resultaron ser menos cooperativos durante la entrevista y presentaron una mayor *no respuesta de ítem* en varias preguntas, como las relativas a renta económica.

- *Incrementar el tamaño de la muestra.* Siempre que no se traduzca en una sobrerrepresentación de los más fáciles de contactar.
- *Sustitución de no respondientes por unidades equiparables.* El remedio más económico y rápido, aunque no el más adecuado. Puede generar más *error de no respuesta* si los «sustitutos» elegidos no son equiparables a quienes reemplazan<sup>38</sup>. La *sustitución*, por la pérdida en representatividad muestral que pudiera suponer, debería ser el último recurso a aplicar, cuando se quiere mantener el tamaño original de la *muestra*<sup>39</sup>. Ahora bien, cuando se lleve a cabo debería especificarse en el informe de la encuesta, para la correcta interpretación de los datos.
- *Ajustes estadísticos mediante la ponderación de la muestra y la imputación.* Su efectividad depende de que la *no respuesta* sea aleatoria en las variables clave de cada investigación, y que las empleadas para la ponderación realmente diferencien a los que responden de los que no lo hacen. Lo que rara vez acontece. Por lo que se aconseja disminuir esta práctica.

La reducción de la *no respuesta de ítem* también supone introducir mejoras en el diseño del cuestionario. Depende de la extensión del cuestionario (si es demasiado largo, cansa al encuestado y éste decide no responderlo por completo), del tipo de preguntas y de su ubicación. El cuestionario incluye preguntas abiertas (más vulnerables a la *no respuesta*) o varias preguntas *filtro*, cuyas rutas no han quedado bien especificadas. Las mejoras en el cuestionario son también imprescindibles para la reducción de la generalidad de los *errores de medición*<sup>40</sup>, como puede verse en el cuadro 2.

Merecen destacarse las mejoras en el cuestionario encaminadas a evitar la presencia de *errores de medición* que se deban a:

- a) La desigual interpretación de términos de la pregunta, por empleo de palabras ambiguas o con distintas connotaciones. De manera especial en los cuestionarios autoadminis-

<sup>38</sup> Es evidente que si el trabajo de campo debe realizarse en dos o tres días, la *sustitución* será mayor que cuando puede dilatarse algunos días más. Ello repercute en la infrarrepresentación en la muestra final de jóvenes, de personas que trabajan, de aquellas con una agenda social más nutrida o se mueven más (Wert, 2002).

<sup>39</sup> En la *Encuesta Social Europea* no se ha admitido ningún tipo de *sustitución*, ni de los casos de «rechazo» ni de los «no contacto». Tampoco la práctica habitual de seleccionar los individuos, en la última etapa de muestreo, según cuotas de sexo y edad. La tasa de respuesta mínima se fijó en el 70%. Si bien éste fue un requisito difícil de cumplir en un gran número de países. España fue el cuarto país con peor tasa de respuesta (53%), sólo superado por Suiza (34%), Italia (44%) y Luxemburgo (46%). En el polo contrario se sitúan Grecia (80%), Polonia (73%) y Finlandia (73%) (Riba y Cuxart, 2004).

<sup>40</sup> Sudman y Bradburn (1974) destacan el diseño del cuestionario como la principal fuente de error en las estimaciones de encuesta. Fowler y Mangione (1990) lo consideran una vía clave en la minimización de los efectos debidos a la entrevista.

## CUADRO 2

## Remedios ante los errores de observación o medición

| Errores de encuesta            | Remedios   |
|--------------------------------|--|
| Por método de encuesta         | <ul style="list-style-type: none"> <li>— Elección razonada del método de encuesta, adecuándose al tema de estudio, su complejidad, la población de interés, el tiempo y los recursos disponibles</li> <li>— Mejoras en el diseño del cuestionario</li> <li>— Mejoras en su aplicación</li> </ul> |
| Por diseño del cuestionario    | <ul style="list-style-type: none"> <li>— Ajustar la formulación de la pregunta y su ubicación al tema y a la población</li> <li>— Precomprobar el cuestionario</li> </ul>  |
| Por entrevistador              | <ul style="list-style-type: none"> <li>— Aumentar el número de entrevistadores</li> <li>— Adecuar su selección a la encuesta y a la población a encuestar</li> <li>— Mejorar su formación</li> <li>— Estandarizar su actuación</li> <li>— Intensificar su supervisión y control</li> </ul>       |
| Por encuestado                 | <ul style="list-style-type: none"> <li>— Mejoras en el diseño del cuestionario</li> <li>— Mejoras en su aplicación</li> </ul>  |
| Por tratamiento de información | <ul style="list-style-type: none"> <li>— Automatización del procedimiento de encuesta (sistemas CADAC o los modos automatizados de grabación mediante escáner, fibra óptica, grabación oral)</li> <li>— Mejorar la formación y el control de los responsables</li> </ul>                         |

trados, en los que un entrevistador no puede aclarar el significado de la pregunta ni de las respuestas.

b) La redacción de la pregunta, su introducción y formulación concretas. La *objetividad* se persigue como requisito de *calidad* de la pregunta. Como dice Payne (1980: 72), «una “buena” pregunta es, entre otras cosas, aquella que no afecta a la respuesta».

c) Las opciones de respuesta: qué alternativas se ofrezcan y en qué orden. En encuestas autoadministradas, o cuando pueden visualizarse las alternativas de respuesta (en una tarjeta u *hoja de respuestas*), se da más el *efecto de primacía*. Las respuestas que figuran al principio son las que reciben un procesamiento cognitivo más profundo (Schuman y Presser, 1981/1996; Bishop *et al.*, 1988; Krosnick y Alwin, 1987). Por el contrario, cuando es el entrevistador quien lee las alternativas de respuesta, las últimas son las que obtienen un procesamiento cognitivo mayor: *efecto de recencia*. Ambos efectos se dan más en personas con un bajo nivel de estudios. Al igual que sucede con el *sesgo de aquiescencia* o tendencia al acuerdo indistintamente de lo que se pregunte (Ross y Mirowsky, 1998). Aunque pueden deberse a problemas de memoria, de recordar las alterna-

tivas de respuesta, parece que es la *saliencia* o relevancia del tema lo más determinante (McClendon, 1991/2002).

d) El orden y el contexto de la pregunta. El *efecto de orden* se produce cuando el encuestado quiere que su respuesta a una segunda pregunta sea consistente con la dada a una anterior. Las preguntas generales, más que las específicas, son más vulnerables, sobre todo cuando estas últimas las anteceden<sup>41</sup>.

e) El tema de la pregunta. Principalmente, cuando ésta pueda inhibir al encuestado de dar una respuesta veraz. Un requisito necesario para la captación de respuestas más sinceras es transmitir al encuestado la sensación de mayor anonimato y confidencialidad<sup>42</sup>. También, el ocultar la intencionalidad de la pregunta; el proporcionar respuestas flexibles o atenuar la gravedad de la pregunta; el redactar la pregunta en un vocabulario familiar al encuestado, en un formato largo (más que corto) y abierto; el aleatorizar las opciones de respuesta o el cargar deliberadamente la pregunta (asumir la conducta y preguntar por su frecuencia u otros detalles).

A diferencia de los otros *sesgos*, el de *deseabilidad social* (el dar la respuesta que se percibe como socialmente deseable) es más habitual en personas de mayor nivel educativo, que perciben más la intencionalidad de la pregunta. También, en sociedades como las hispanas, más afectadas por el mundo de las apariencias, por el «qué dirán». En cambio, los jóvenes se muestran menos susceptibles a este sesgo<sup>43</sup>.

Estos y otros *errores de medición* requieren, para su detección y reducción, la previa comprobación del cuestionario. Pero siempre que el *pretest* se ajuste al diseño de la encuesta, en cada una de sus fases, a las condiciones «reales» en las que finalmente se llevará a cabo. Tal es la importancia que se concede al *pretest* que Sudman y Bradburn (1982/1987) aconsejan no realizar la encuesta si se carece de recursos para la prueba piloto del cuestionario. En su opinión, habría que hacer, como mínimo, cuatro pruebas o revisiones de los cuestionarios antes de su lanzamiento definitivo. Converse y Presser (1994) las reducen a

<sup>41</sup> Ya en 1939, Sayre comprobó que las preguntas precedentes pueden influir en las respuestas a preguntas posteriores. Varios de sus hallazgos fueron después confirmados por Cantril (1944) y otros autores, como Schuman y Presser (1981/1996). Estos últimos constatan, por ejemplo, que una pregunta genérica sobre la felicidad obtiene un porcentaje mayor de respuestas favorables cuando es preguntada primero (52%) que después de una específica sobre la felicidad conyugal (38%).

<sup>42</sup> Ayuda el uso de *hojas de respuestas* en la encuesta *cara a cara*. Aquilino (1994), Sudman y Bradburn (1974) o Johnson (1989) lo constatan en sus estudios de admisión de conductas no deseables.

<sup>43</sup> Esto último es demostrado por Scott (1997), y la incidencia del contexto sociocultural por Johnson y colaboradores (1997), en su estudio comparativo de las sociedades mejicana y anglosajona. Por su parte, Krysan (1998) muestra la mayor incidencia de este sesgo en personas de mayor nivel educativo, que tienen menos interés en parecer «racistas», como también se muestra en el estudio de Cea D'Ancona (2004a).

dos; Oppenheim (1966) incrementa el número a ocho. De Lamater y McKinney (1982) censuran la práctica estándar de pretestar el cuestionario en muestras de conveniencia de estudiantes, colegas o amigos, a menos que sean la población de estudio.

Aparte del *pretest* existen otros procedimientos alternativos de comprobación de la *validez de medición* de la encuesta: experimentos *split-ballot*, estudios de *contrastación* de los datos de encuesta con los datos de registros administrativos, la *observación* del ajuste entre el diseño de la encuesta y su realización, la *evaluación* de los cuestionarios (por expertos) y los estudios de *grupo de criterio* (pasar el cuestionario a un grupo de personas de las que se sabe, con un cierto grado de certeza, que poseen abundancia o deficiencia del concepto que se mide). Todo lo cual ayuda a detectar problemas de *validez* y de *fiabilidad* en la encuesta.

De ellos se desprenderán actuaciones dirigidas a la mejora y aplicación del cuestionario. Eso sí, cualquier modificación que se haga requiere su precomprobación y, en caso de no conseguir la mejora esperada, proceder a su revisión, modificación y consiguiente precomprobación. Como indica Sheatsley (1983: 228), «un error común es escribir enteramente nuevas preguntas, cambiar la redacción de muchas otras o asumir que se ha cortado diez minutos del tiempo de entrevista, haciendo ciertas eliminaciones, y luego enviar la versión revisada al campo sin probarla». Lamentablemente, esta prueba del cuestionario continua no es usual en la práctica de la *encuesta*, por el incremento que supone en los costes del estudio, en tiempo y dinero. No obstante, sería deseable en encuestas pioneras, que no siguen el hacer de estudios precedentes.

### 3. BALANCE FINAL: LOS VIEJOS NUEVOS RETOS

En las páginas precedentes se ha ido exponiendo la relación de errores y de aspectos a considerar en la evaluación de la *calidad* de la encuesta. El lector habrá deducido la interrelación entre ellos y cómo una mala decisión o actuación inicial afecta negativamente a esa fase en el desarrollo de la encuesta y a posteriores, deteriorando su *calidad*.

Los modos informatizados de encuesta cada vez van adquiriendo un mayor protagonismo en la investigación. Principalmente, por su contribución en la disminución de los costes económicos y del período de tiempo necesario para disponer de los datos (las fases afectadas son las de recogida de información, grabación y edición) y de los *errores* en la administración del cuestionario. No sólo propician un mayor control de la actuación de los entrevistadores, al grabarse información relativa a la ejecución de la entrevista (hora de comienzo y de conclusión, la secuencia seguida en su realización). También se logra la re-

ducción de errores en la grabación de las respuestas (de registro o de inconsistencia) que pueden resolverse durante la entrevista. Y de los errores debidos a saltos de preguntas porque la *ruta* a seguir (en la formulación de las preguntas a cada encuestado) queda marcada automáticamente.

Además, pueden contribuir a la disminución de los efectos de *recencia*, de *primacia* y demás referidos al orden de las respuestas y de las preguntas (al poderse programar el cuestionario para ofrecer a cada encuestado un orden de preguntas y de respuestas diferente). Incluso parecen provocar una menor inhibición en los entrevistados en preguntas sensibles o delicadas (De Leeuw *et al.*, 1995/2002). Lo que puede rebajar el *sesgo de deseabilidad social*, al transmitir al encuestado un sentimiento de mayor privacidad. También incrementan las opciones de ayudas visuales, que favorecen la comprensión de las preguntas. Y, en general, transmiten una imagen de mayor profesionalidad, que afecta positivamente a la participación en la encuesta. Aunque en personas no familiarizadas con el uso de ordenadores pueden provocar el efecto contrario, más si la encuesta es autoadministrada.

Pero no son la panacea. La informatización del procedimiento de encuesta pone remedio a ciertas deficiencias, si bien quedan muchos flancos por resolver. Dependen no sólo de condicionantes económicos y temporales. También del grado de formación y de destreza de los responsables de cada fase de la encuesta, sin olvidar la existencia de «controles de calidad» en cada una de ellas. Cada vez es más exigible un mayor «rigor» en cualquier decisión y actuación. De modo que si se carece de los recursos y de la capacidad suficientes para llevar a efecto con éxito una encuesta, mejor no hacerla. Óptese por otra estrategia metodológica, por bien de su investigación y de la encuesta en general.

Asimismo, se demanda una mayor precisión en la interpretación de los datos de encuesta. La *ficha técnica*, exigida como garante de la «calidad» de los datos que la encuesta ofrece, muestra ser claramente insuficiente en su calibración, por circunscribirse a sólo unos pocos datos técnicos en su realización (relacionados con el error muestral). La correcta interpretación de los datos de encuesta, y su consiguiente *credibilidad*, precisan de la especificación de los «límites» de la investigación<sup>44</sup>. Junto con el detalle de las actuaciones preventivas y paliativas emprendidas para su corrección. Sólo así se conseguirá avanzar en la *senda tortuosa de la calidad de la encuesta*.

<sup>44</sup> Como hizo SORA (2001) en su informe del *Eurobarómetro 53: Attitudes towards minority groups in the European Union*, de acceso gratuito en la red (<http://www.europa.eu.int/comm/dg10/epol>); ejemplo de reconocimiento de todos los errores de observación y de medición que detectaron en su realización.

## BIBLIOGRAFÍA

- ÁLVAREZ GONZÁLEZ, R. (2004): *Datos para la evaluación de la calidad de la encuesta*, ponencia presentada al VIII Congreso Español de Sociología (Alicante, 23-25 de septiembre).
- AQUILINO, W. S. (1994): «Interview mode effects in surveys of drug and alcohol use. A field experiment», *Public Opinion Quarterly*, 58: 210-240.
- ARMSTRONG, J. S. (1975): «Monetary incentives in mail surveys», *Public Opinion Quarterly*, 39: 223-250.
- BERK, M. L.; MATHIOWETZ, N. A.; WARD, E. P., y WHITE, A. A. (1987): «The effect of prepaid and promised incentives: results of a controlled experiment», *Journal of Official Statistics*, 3 (4): 449-457.
- BIEMER, P. P., y LYBERG, L. E. (2003): *Introduction to survey quality*, John Wiley & Sons, New Jersey.
- BISHOP, G. F.; HIPPLER, H. J.; SCHWARZ, N., y STRACK, F. (1988): «A comparison of response effects in self-administered and telephone surveys», en R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II y J. Waksberg, *Telephone survey methodology*, John Wiley & Sons, New York, pp. 321-340.
- BISHOP, G. F.; TUCHFARBER, A. J., y OLDENDICK, R. W. (1986): «Opinions on fictitious issues: the pressure to answer survey questions», *Public Opinion Quarterly*, 50: 240-250.
- BRADBURN, N. M.; SUDMAN, S., y colaboradores (1979): *Improving interview method and questionnaire design: response effects to threatening questions in survey research*, Jossey-Bass, San Francisco.
- BRADBURN, N. M.; SUDMAN, S.; BLAIR, E., y STOCKING, C. (1978): «Question threat and response bias», *Public Opinion Quarterly*, 42: 221-234.
- BRENNAN, M., y HOEK, J. (1992): «The behavior of respondents, nonrespondents and refusers across mail surveys», *Public Opinion Quarterly*, 56: 530-535.
- BRUNNER, G. A., y CARROLL, S. J. (1969): «The effect of prior notification on the refusal rate in fixed address surveys», *Journal of Marketing Research*, 9: 42-44.
- CAMACHO, C.; PRADO, C.; ROMERO, B., y VALERA, S. (2000): «¿Cómo hacen encuestas los encuestadores?», *Metodología de Encuestas*, 2 (1): 117-133.
- CAMPANELLI, P.; THOMSON, K.; MOON, N., y STAPLES, T. (1997): «The quality of occupational coding in the United Kingdom», en L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz y D. Trewin, *Survey measurement and process quality*, John Wiley & Sons Inc., New York, pp. 437-455.
- CAMPBELL, D. T., y FISKE, D. W. (1959): «Convergent and discriminant validation by the multitrait-multimethod matrix», *Psychological Bulletin*, 56: 81-105.
- CAMPBELL, D. T., y STANLEY, J. (1970): *Diseños experimentales y cuasiexperimentales en la investigación social*, Amorrortu, Buenos Aires.
- CANTRIL, H. (1944): *Gauging public opinion*, Princeton University Press, Princeton.
- CEA D'ANCONA, M.<sup>a</sup> A. (2002): «La medición de las actitudes ante la inmigración: evaluación de los indicadores tradicionales de "racismo"», *Revista Española de Investigaciones Sociológicas*, 99: 87-111.
- (2004a): *La activación de la xenofobia en España. ¿Qué miden las encuestas?*, Centro de Investigaciones Sociológicas, Colección Monografías, n.º 211, Madrid.
- (2004b): *Métodos de encuesta. Teoría y práctica, errores y mejora*, Síntesis, Madrid.



- CIALDINI, R. B. (1984): *Influence: the new psychology of modern persuasion*, Harper Collins, New York.
- COLLINS, M.; SYKES, W.; WILSON, P., y BLACKSHAW, N. (1988): «Nonresponse: the U.K. experience», en R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II y J. Waksberg, *Telephone survey methodology*, John Wiley & Sons, New York, pp. 213-231.
- CONVERSE, J. M., y PRESSER, S. (1994): «Survey questions: handcrafting the standardized questionnaire», en M. S. Lewis-Beck, *Research Practice*, Sage, California, pp. 89-162.
- COOK, T. D., y CAMPBELL, D. T. (1977): «The design and conduct of quasi-experiments and true experiments in field settings», en H. M. Dunette, *Handbook of industrial psychology*, Rand McNally, Chicago.
- COUPER, M. P., y HANSEN, S. E. (2002): «Computer-assisted interviewing», en J. F. Gubrium y J. A. Holstein, *Handbook of interview research. Context & Methods*, Sage, Thousand Oaks, California, pp. 557-575.
- CRONBACH, L. J. (1951): «Coefficient alpha and the internal structure of tests», *Psychometrika*, 16: 297-334.
- CURTIN, R.; PRESSER, S., y SINGER, E. (2000): «The effects of response rate changes on the index of consumer sentiment», *Public Opinion Quarterly*, 64: 413-428.
- DAVIS, J. A., y SMITH, T. W. (1992): *The NORC General Social Survey. A user's guide*, Sage, Newbury Park.
- DE LAMATER, J., y MCKINNEY, K. (1982): «Response-effects of question content», en W. Dijkstra y J. van der Zouwen, *Response behaviour in the survey-interview*, Academic Press, New York, pp. 13-48.
- DE LEEUW, E. D.; HOX, J. J., y SNIJKERS, G. (2002): «The effect of computer-assisted interviewing on data quality: a review», en D. De Vaus, *Social surveys*, Sage, London, pp. 170-191 (e. o. 1995: *Journal of the Market Research Society*, 37: 325-344).
- DÍAZ DE RADA IGÚZQUIZA, V. (2000): *Problemas originados por la no respuesta en la investigación social: definición, control y tratamiento*, Universidad Pública de Navarra, Navarra.
- (2001): «Problemas de cobertura en la encuesta telefónica», *Revista Española de Investigaciones Sociológicas*, 93: 133-164.
- DEMING, W. E. (1944): «On errors in surveys», *American Sociological Review*, 9: 359-369.
- DILLMAN, D. A. (1978): *Mail and telephone surveys: the total design method*, John Wiley & Sons, New York.
- (1991): «The design and administration of mail surveys», *Annual Review of Sociology*, 17: 225-249.
- (2000): *Mail and Internet surveys: the tailored design method*, Wiley, New York.
- DILLMAN, D. A.; GALLEGOS, J. G., y FREY, J. H. (1976): «Reducing refusal rates for telephone interviews», *Public Opinion Quarterly*, 40: 66-78.
- FARIA, A. J.; DICKINSON, J. R., y FILIPIC, T. V. (1990): «The effect of telephone versus letter prenotification on mail survey response rate», *Journal of the Market Research Society*, 32: 551-568.
- FELDMAN, J. J.; HYMAN, H., y HART, C. W. (2002): «A field study of interviewer effects on the quality of survey data», en D. De Vaus, *Social surveys*, Sage, London, pp. 118-141 (e. o. 1951-1952: *Public Opinion Quarterly*, 15: 734-761).
- FERNÁNDEZ ESQUINAS, M. (2003): «Criterios de calidad en la investigación social: la producción de datos sociales», *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 6: 47-77.
- FOWLER, F. J., Jr. (1993): *Survey research methods*, Sage, Newbury Park, California.

- FOWLER, F. J., Jr., y MANGIONE, T. W. (1990): *Standardized survey interviewing: minimizing interviewer-related error*, Sage, Newbury Park, California.
- FOX, R. J.; CRASK, M. R., y KIM, J. (1988): «Mail survey response rate: a meta-analysis of selected techniques for inducing response», *Public Opinion Quarterly*, 52: 467-491.
- FREY, J. H., y OISHI, S. M. (1995): *How to conduct interviews by telephone and in person*, Sage, Thousand Oaks, California.
- GELB, B. D. (1975): «Incentives to increase survey returns: social class considerations», *Journal of Marketing Research*, 12: 107-109.
- GILLIN, J. L. (1915): «The social survey and its further development», *Journal of the American Statistical Association*, 14: 603-610.
- GOODSTADT, M. S.; CHUNG, L.; KRONITZ, R., y COOK, G. (1977): «Mail survey response rates: their manipulation and impact», *Journal of Marketing Research*, 14: 391-395.
- GROVES, R. M. (1989): *Survey errors and survey costs*, John Wiley & Sons, New York.
- GROVES, R. M.; CIALDINI, R. B., y COUPER, M. P. (1992): «Understanding the decision to participate in a survey», *Public Opinion Quarterly*, 56: 475-495.
- GROVES, R. M., y COUPER, M. P. (1998): *Nonresponse in the Household interview surveys*, Wiley-Interscience, New York.
- GROVES, R. M., y FULTZ, N. H. (1985): «Gender effects among telephone interviewers in a survey of economic attitudes», *Sociological Methods and Research*, 14 (1): 31-52.
- GROVES, R. M., y LYBERG, L. E. (1988): «An overview of nonresponse issues in telephone surveys», en R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II y J. Waksberg, *Telephone survey methodology*, John Wiley & Sons, New York, pp. 191-211.
- GROVES, R. M., y MAGILAVY, L. J. (1981): «Increasing response rates to telephone surveys: a door in the face for foot-in-the door?», *Public Opinion Quarterly*, 45: 346-358.
- GROVES, R. M.; SINGER, E., y CORNING, A. (2000): «Leverage-saliency theory of survey participation. Description and an illustration», *Public Opinion Quarterly*, 64: 299-308.
- GOULDNER, A. W. (1960): «The norm of reciprocity: a preliminary statement», *American Sociological Review*, 25 (2): 161-178.
- GOYDER, J. C. (1987): *The silent minority. Nonrespondents on sample surveys*, Polity Press, Blackwell, Cambridge.
- HANSEN, R. A. (1980): «A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey response behaviour», *Journal of Marketing Research*, 17: 77-83.
- HANSEN, M. H.; HURWITZ, W. N., y MADOW, W. G. (1953): *Sample survey methods*, Wiley, New York.
- HEBERLEIN, T. A., y BAUMGARTNER, R. (1978): «Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature», *American Sociological Review*, 43: 447-462.
- HENERSON, M. E.; MORRIS, L. L., y FITZ-GIBBON, C. T. (1987): *How to measure attitudes*, Sage, Newbury Park, California.
- HOBSON, A. (1916): «The use of the correspondence method in original research», *Journal of the American Statistical Association*, 15: 210-218.

- HOUSE, J. S.; GERBER, W., y McMICHAEL, A. J. (1977): «Increasing mail questionnaire response: a controlled replication and extension», *Public Opinion Quarterly*, 41: 95-99.
- HOX, J. J., y DE LEEUW, E. D. (1994): «A comparison of nonresponse in mail, telephone and face to face surveys», *Quality & Quantity*, 28: 329-344.
- (2002): «A comparison of nonresponse in mail, telephone and face to face surveys», en D. De Vaus, *Social surveys*, Sage, London, pp. 157-171.
- HUBBARD, R., y LITTLE, E. (1988): «Promised contribution to charity and mail survey responses», *Public Opinion Quarterly*, 52: 223-230.
- HYMAN, H., *et al.* (1954): *Interviewing in social research*, University Press, Chicago.
- JAMES, J. M., y BOLSTEIN, R. (1990): «The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys», *Public Opinion Quarterly*, 54: 346-361.
- JOHNSON, T. P., *et al.* (1989): «Obtaining reports of sensitive behaviour: a comparison of substance use reports from telephone and face-to-face interviews», *Social Science Quarterly*, 70: 174-183.
- JOHNSON, T. P.; O'ROURKE, D.; CHAVEZ, N.; SUDMAN, S.; WARNECKE, R.; LACEY, L., y HORM, J. (1997): «Social cognition and responses to survey questions among culturally diverse populations», en L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz y D. Trewin, *Survey measurement and process quality*, John Wiley & Sons Inc., New York, pp. 87-114.
- KAHN, R. L., y CANNELL, C. F. (1957): *The dynamics of interviewing: theory, techniques and cases*, John Wiley and Sons, New York.
- KALTON, G.; COLLINS, M., y BROOK, L. (1978): «Experiments in wording opinion questions», *Applied Statistics*, 27: 149-161.
- KANE, E. W., y MACAULAY, L. F. (1993): «Interviewer gender and gender attitudes», *Public Opinion Quarterly*, 57: 1-28.
- KATZ, D. (1942): «Do interviewers bias polls?», *Public Opinion Quarterly*, 6: 248-268.
- KEETER, S. (1995): «Estimating telephone noncoverage bias with a telephone survey», *Public Opinion Quarterly*, 59: 196-217.
- KIAER, A. N. (1897): *Representative method of statistical surveys*, Kristinia, Oslo.
- KISH, L. (1965/1995): *Survey sampling*, John Wiley & Sons, New York.
- KROSNICK, J. A., y ALWIN, D. F. (1987): «An evaluation of a cognitive theory of response-order effects in survey measurements», *Public Opinion Quarterly*, 51: 201-219.
- KRYSAN, M. (1998): «Privacy and the expression of white racial attitudes. A comparison across three contexts», *Public Opinion Quarterly*, 62: 506-544.
- KUDER, G. F., y RICHARD, M. W. (1937): «The theory of the estimation of test reliability», *Psychometrika*, 2: 135-138.
- LAVRAKAS, P. J. (1993): *Telephone survey methods: sampling, selection and supervision*, Sage, Newbury Park (e. o. 1987).
- (1998): «Methods of sampling and interviewing in telephone surveys», en L. Bickman y D. J. Rog, *Handbook of applied social research methods*, Sage, Thousand Oaks, California, pp. 429-472.

- LOSCH, M. E.; MAITLAND, A.; LUTZ, G.; MARIOLIS, P., y GLEASON, S. (2002): «The effect of time of year of data collection on sample efficiency. An analysis of behavioural risk factor surveillance survey data», *Public Opinion Quarterly*, 66: 594-607.
- MAHALANOBIS, P. C. (1946): «Recent experiments in statistical sampling in the Indian Statistical Institute», *Journal of the Royal Statistical Society*, 19: 325-378.
- MANGIONE, T. W. (1998): «Mail surveys», en L. Bickman y D. J. Rog, *Handbook of applied social research methods*, Sage, Thousand Oaks, California, pp. 399-428.
- McCLENDON, M. J. (2002): «Acquiescence and recency response-order effects in interview surveys», en D. De Vaus, *Social surveys*, Sage, London, pp. 84-117 [e. o. 1991: *Sociological Methods and Research*, 20 (1): 60-103].
- NEYMAN, J. (1934): «On the different aspects of the representative method: the method of stratified sampling and the method of purposive selection», *Journal of the Royal Statistical Society*, 97: 558-606.
- NÚÑEZ VILLUENDAS, A. (2004): *Incidencias de la entrevista personal en la investigación por encuesta*, comunicación presentada al VIII Congreso Español de Sociología (Alicante, 23-25 de septiembre).
- OKSENBERG, L., y CANNELL, Ch. (1988): «Effects of interviewer vocal characteristics on nonresponse», en R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II y J. Waksberg, *Telephone survey methodology*, John Wiley & Sons, New York, pp. 257-269.
- OPPENHEIM, A. (1966): *Questionnaire design and attitude measurement*, Heinemann, London.
- ORNE, M. T. (1969): «Demand characteristics and the concept of quasi-controls», en R. Rosenthal y R. L. Rosnow, *Artifacts in behavioral research*, Academic Press, New York, pp. 143-179.
- PAYNE, S. L. (1980): *The art of asking questions*, Princeton University Press, Princeton, New Jersey (e.o. 1951).
- REICHARDT, Ch. S., y COOK, T. D. (1979): «Beyond qualitative versus quantitative methods», en T. D. Cook y Ch. Reichardt, *Qualitative and quantitative methods in social research*, Sage, Beverly Hills.
- RIBA, C., y CUXART, A. (2004): *Aspectos metodológicos de la Encuesta Social Europea*, ponencia presentada al VIII Congreso Español de Sociología (Alicante, 23-25 de septiembre).
- RICE, S. A. (1929): «Contagious bias in the interview: a methodological note», *American Journal of Sociology*, 35: 420-423.
- ROSS, C. E., y MIROWSKY, J. (1998): «The worst place and the best face», *Social Forces*, 62: 529-536.
- SÁNCHEZ CARRIÓN, J. J. (1996): «La calidad de la encuesta», *Papers*, 48: 127-146.
- (2000): *La bondad de la encuesta: el caso de la no respuesta*, Alianza, Madrid.
- SAYRE, J. (1939): «A comparison of three indexes of attitudes towards radio advertising», *Journal of Applied Psychology*, 23: 23-33.
- SCHAEFFER, N. C., y MAYNARD, D. W. (2002): «Standardization and interaction in the survey interview», en J. F. Gubrium y J. A. Holstein, *Handbook of interview research. Context & Methods*, Sage, Thousand Oaks, California, pp. 577-601.
- SCHUMAN, H., y PRESSER, S. (1981): *Questions & answers in attitude surveys. Experiments on question form, wording and context*, Academic Press, New York / (1996) Sage, Thousand Oaks, California.
- SCOTT, C. (1961): «Research on mail surveys», *Journal of Royal Statistical Society*, 124: 143-205.

- SCOTT, J. (1997): «Children as respondents: methods of improving data quality», en L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz y D. Trewin, *Survey measurement and process quality*, John Wiley & Sons Inc., New York, pp. 331-351.
- SHEATSLEY, P. B. (1983): «Questionnaire construction», en P. H. Rossi, J. D. Wrigth y A. B. Anderson, *Handbook of survey research*, Academic Press Inc., Orlando, pp. 195-230.
- SINGER, E. (1978): «Informed consent: consequences for response rate and response quality in social surveys», *American Sociological Review*, 43: 144-162.
- (2002): «Informed consent and survey response: a summary of the empirical literature», en D. De Vaus, *Social surveys*, Sage, London, pp. 280-295 [e. o. 1993: *Journal of Official Statistics*, 9 (2): 361-375].
- SINGER, E., y KÖHNKE-AGUIRRE, L. (1979): «Interviewer expectation effects: a replication and extension», *Public Opinion Quarterly*, 43: 245-260.
- SINGER, E.; VAN HOEWYK, J., y MAHER, M. P. (1998): «Does the payment of incentives create expectation effects?», *Public Opinion Quarterly*, 62: 152-164.
- (2000): «Experiments with incentives in telephone surveys», *Public Opinion Quarterly*, 64: 171-188.
- SINGER, E.; VON THURN, D. R., y MILLER, E. R. (1995): «Confidentiality assurances and response: a quantitative review of the experimental literature», *Public Opinion Quarterly*, 59: 66-77.
- SINGLETON, R. A., Jr., y STRAITS, B. C. (2002): «Survey interviewing», en J. F. Gubrium y J. A. Holstein, *Handbook of interview research. Context & Methods*, Sage, Thousand Oaks, CA, pp. 59-82.
- SORA (Thalhammer, E.; Enzenhofer, E.; Salfinger, B., y Ogris, G.) (2001): *Attitudes towards minority groups in the European Union. A special analysis of the Eurobarometer 2000 survey*, European Monitoring Centre on Racism and Xenophobia, Viena.
- SUDMAN, S., y BRADBURN, N. M. (1974): *Response effects on surveys: a review and synthesis*, Aldine, Chicago.
- (1982/1987): *Asking questions: a practical guide to questionnaire design*, Jossey-Bass, San Francisco.
- TRAUGOTT, M. W. (1987): «The importance of persistence in respondent selection for preelection surveys», *Public Opinion Quarterly*, 51: 48-57.
- TRAUGOTT, M. W.; GROVES, R. M., y LEPKOWSKI, J. (1987): «Using dual frame designs to reduce nonresponse in telephone surveys», *Public Opinion Quarterly*, 51: 522-539.
- TRUJILLO CARMONA, M., y SERRANO DEL ROSAL, R. (2003): «Análisis de la incidencia de las llamadas en las encuestas telefónicas», *Metodología de Encuestas*, 5 (1): 31-44.
- VAN KAMMEN, W. B., y STOUTHAMER-LOEBER, M. (1998): «Practical aspects of interview data collection and data management», en L. Bickman y D. J. Rog, *Handbook of applied social research methods*, Sage, Thousand Oaks, California, pp. 375-398.
- VIGDERHOUS, G. (1981): «Scheduling phone interviews: a study of seasonal patterns», *Public Opinion Quarterly*, 45: 250-259.
- WARRINER, K.; GOYDER, J.; GJERRSEN, H.; HOHNER, P., y McSPURREN, K. (1996): «Charities, no; lotteries, no; cash, yes», *Public Opinion Quarterly*, 60: 542-562.
- WERT, J. I. (2002): «No le contéis a mi madre que hago encuestas políticas. Ella cree aún que soy pianista en un burdel», *Revista Española de Investigaciones Sociológicas*, 99: 237-245.
- WOTRUBA, T. R. (1966): «Monetary inducements and mail questionnaire response», *Journal of Marketing Research*, 3: 393-400.