

Pretesting survey instruments: An overview of cognitive methods

Debbie Collins

National Centre for Social Research, London, UK (E-mail: d.collins@natcen.ac.uk)

Accepted in revised form 24 January 2001

Abstract

This article puts forward the case that survey questionnaires, which are a type of measuring instrument, can and should be tested to ensure they meet their purpose. Traditionally survey researchers have been pre-occupied with 'standardising' data collection instruments and procedures such as question wording and have assumed that experience in questionnaire design, coupled with pilot testing of questionnaires, will then ensure valid and reliable results. However, implicit in the notion of standardisation are the assumptions that respondents are able to understand the questions being asked, that questions are understood in the same way by all respondents, and that respondents are willing and able to answer such questions. The development of cognitive question testing methods has provided social researchers with a number of theories and tools to test these assumptions, and to develop better survey instruments and questionnaires. This paper describes some of these theories and tools, and argues that cognitive testing should be a standard part of the development process of any survey instrument.

Key words: Cognitive aspects of survey methodology, Cognitive interviews, Data quality

Introduction

Whether we are social researchers or epidemiologists, designing surveys or clinical trials, we aim to provide results that are valid, reliable, sensitive, unbiased and complete. In other words we want to be certain that our questions measure the concepts or behaviours we want them to measure, that the data produced represent 'true' values for these measures and do not contain too much random variability, that our questions are sensitive enough to measure important real differences or changes, and that our study covers all the dimensions of the topic under investigation [1].

These goals may sound straightforward enough but their implementation can be far from easy. Often studies are designed within constraints on the amount of money and time available for instrument testing and development. In light of the aims and priorities of the study decisions are made about the level of accuracy of the data required.

Such decisions often involve considerations of sample size and design, and mode of data collection. Whilst all these issues are important and are commonly used as measures of survey quality we should not forget the role of the data collection instrument – often a questionnaire or diary – on data quality. And it is on the design of data collection instruments that yield reliable, valid, sensitive, unbiased and complete data that the rest of this paper focuses.

Ways of thinking about measurement error

Implicit in any form of measurement is the notion of a standard procedure or set of tools that are used to obtain data. In survey research usually the standard data collection tool is the questionnaire and the measurements obtained are the respondent's answers to survey questions. A further stage of standardisation is required where the

questionnaire is to be administered by interviewers, to ensure that all interviewers administer it in the same way. By applying the ‘standard’ we hope to ensure that observed differences are in fact real differences and not an artefact of differences in the way the data were collected [2]. However this scientific model of standardisation assumes that:

- all respondents understand the questions in a consistent way;
- the questions are asking for information that respondents have and can retrieve;
- the wording of questions provides respondents with all the necessary information they require to be able to answer them in the way required by the researcher; and
- where interviewers are being used, they always read the questions as worded.

In fact survey researchers have known for many years that despite their attempts to standardise data collection tools and procedures measurement deviations and errors still occur. Traditionally these errors have been classified into two broad categories, those connected with survey questions and those connected with survey interviewers [2]. However, more recently there has been a shift in emphasis, from viewing errors as being the product of either the questionnaire or the interviewer, to being related to the nature of the tasks the actors in a survey interview have to perform [3]. This

task-focused classification is useful in helping us to understand the potential sources of measurement error as it focuses on the specific components of the question-and-answer process. For example, the observation that respondents cannot answer a particular question does not help in identifying the reason why it cannot be answered, and thus finding a solution to the problem is more difficult. The task-focused model on the other hand, would help to identify the cause of the problem by enabling the researcher to identify whether the problem is one of comprehension, processing or communication (see Table 1).

However the task-focused classification of measurement errors proposed by Oksenberg et al. [3] is – in and of itself – of little use unless we can operationalise it. We need tools and methods that enable us to test questions and identify potential sources of measurement error, but we also would like tools that will enable us to quantify the size and direction of any measurement error.

Ideally we would like three kinds of evidence to evaluate the performance of survey questions:

Statistical – identifying the specific effect of question measurement error on survey estimates.

Direct study of the question-and-answer process – identifying how and where the question fails to achieve its measurement purpose.

Table 1. Components of measurement error

Traditional model	Task-focused model
(1) Problems with survey questions: <ul style="list-style-type: none"> – that are misunderstood – that cannot be answered, either at all or accurately – that respondents will not answer 	(1) Comprehension problems resulting from: <ul style="list-style-type: none"> – use of vocabulary – complex sentence structure – not understanding the nature of the task and the rules about how to respond
(2) Problems with survey interviewers: <ul style="list-style-type: none"> – do not read the questions as worded – probe directly – bias answers as a result of the way interviewers relate to respondents (for example, differences in ethnicity, age, social class, gender) – record answers inaccurately 	(2) Validity problems resulting from: <ul style="list-style-type: none"> – respondents interpreting the same question in different ways, or – in the same way but not in the way the researcher intended
	(3) Processing difficulties: <ul style="list-style-type: none"> – respondents may be unwilling or unable to retrieve the information necessary to answer the question
	(4) Pronunciation or communication difficulties: <ul style="list-style-type: none"> – these may affect both interviewers and respondents

Experimental – identifying whether proposed changes to question forms actually improve data quality.

Cognitive testing of survey questions addresses the second type of evidence and this will be the focus of the rest of this paper. However before describing the theoretical and practical implementation of cognitive methods, we need to take a step back and consider why we need to use such methods to pre-test data collection instruments such as questionnaires.

Why we need to pre-test data collection instruments

Respondents generally seem able and willing to answer survey questions. People want to be polite and so will often answer questions even though they may not understand what the question is asking of them (for example, see Ref. [4]). They will often misunderstand words or concepts, sometimes in different ways, without even realising it. Even though these fundamental communication difficulties may exist, respondents may still provide, what appear to be on the surface, legitimate answers to survey questions by viewing the survey question-and-answer process as following the same rules as day-to-day conversation [5, 6]. They will ‘satisfice’ by dodging the hard work of retrieving information from memory, performing computations or using rating scales, by giving an easily available answer they think will satisfy [7, 8]. Thus the key issue for the researcher is how to check for misunderstandings, incomplete concept coverage, inconsistent interpretations, satisficing, context effects and so on.

Cognitive methods

These methods, derived from social and cognitive psychology, enable us to explore the processes by which respondents answer survey questions, and the factors which influence the answers they provide. Pre-testing questions, particularly pre-testing questions in their questionnaire context, enables us to establish whether:

- respondents can understand the question concept or task,
- they do so in a consistent way, and,
- in a way the researcher intended.

Traditional ‘rehearsal’ piloting alone will not provide evidence of all these types of problems because it usually involves testing the whole survey process. In a pilot the questionnaire is often examined in terms of length, flow, salience, ease of administration and response and acceptability to respondents. Whilst pilots may detect overt problems that disrupt the response elicitation process they often do not provide evidence of causes, nor do they provide evidence of covert problems. Thus we need to use question-testing methods derived from social and cognitive psychology, which will help us to systematically look at the question-and-answer process.

Origin and development of cognitive methods – an overview

Researchers have long been concerned with question wording issues (for example see Refs. [9–13]). Yet it has only been in the past 25 years or so that survey questions and instruments have started to be evaluated in a more detailed and systematic way using theories and methods derived from cognitive and social psychology. The development of the Cognitive Aspects of Survey Methodology (CASM) movement, which grew out of two meetings – one held in the United States in June 1983 [14] and the other held in Germany in July 1984 [15] – has pushed to the fore the importance of respondent task analysis and measurement error. (For a more detailed history of the development of the CASM movement see Tanur [16] and Aborn [17]). The dual development of theories of survey response and cognitive methods has greatly added to our understanding of the sources of measurement error.

Theories of survey response

Before describing some of the cognitive methods used to test survey instruments, it is important to review their theoretical underpinnings. The question-and-answer model, derived from cognitive psychology, is a useful and commonly cited representation of how respondents answer survey questions. In its simplest form the model suggests there are four actions that respondents have to complete

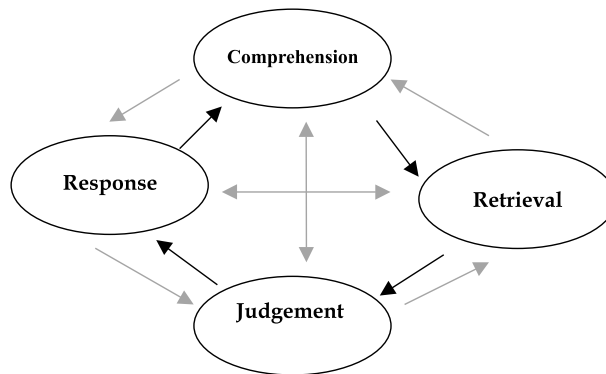


Figure 1. Elaborated question-and-answer model.

in order to answer a question: they must comprehend the question, retrieve the necessary information from long-term memory, make a judgement about the information needed to answer the question, and respond to the question [18].

In reality the question-and-answer process is probably not linear but rather involves numerous iterations of and interactions between the different phases, as shown in Figure 1. For example, respondents may make judgements about the level of detail needed to answer a survey question based on how difficult it is to retrieve the information required and/or by the way in which answers are to be reported or the answer categories provided. Is the question asking for the exact number of occasions I visited my doctor in the past 6 months, or is it asking for an indication of frequency – none, between 1 and 5, 6 or more? Will I appear out of the ordinary if I give the answer I am minded to give? [6, 19].

Let us look at each of the question-and-answer processes in a little more detail.

Comprehension

A key issue for comprehension is whether the respondent understands the question in the same way as the researcher intended. This is important because if the respondent interprets the question in a different way to what was intended by the researcher, conclusions drawn from the respondent's answer may be flawed. Worse still, if different respondents interpret the question in different ways from each other, and from what the researcher

intended, comparisons between respondents' answers will not be valid.

Clearly the goal of the researcher is to design a question that can be understood by all respondents, in the same way, and in a way the researcher intended. However this is more difficult than it first might appear because the meaning of a question or utterance has two components – literal and intended. Literally understanding the words is not sufficient to be able to answer the question. For example, in asking a question about abdominal pain it is important to know whether respondents know what the term 'abdominal' means. This is the literal meaning of the question. However respondents may know what the words mean but may interpret what the purpose of the question is in different ways. For example, what should I report, how much detail is required? These decisions will be influenced by the context in which the question is asked and will draw upon the principles that govern the conduct of daily conversation. This is because we often draw upon our stock of background information and knowledge in interpreting text, which means we will often fill in gaps, add details and make inferences based on our background stock of knowledge and what the survey interview requires [5, 20–22].

Retrieval of information

Having comprehended the question the respondent then (usually) has to retrieve the relevant information from long-term memory, be it factual or attitudinal. In the case of factual information – either current or historical – a number of factors

may affect the retrieval process. Firstly, if the retrieval context is different to the original encoding context the respondent may not be able to recognise that the event took place or be able to recall the correct event [23]. Secondly, the rarer or more distinctive an event is the more likely respondents are to remember it. Consequently, commonly occurring types of event will be harder to distinguish and recall individually [24]. Moreover, over time respondents are likely to have experienced more similar events, so that rare or more distinctive events will become scarcer. This means that accurate recall of many events will become more difficult because there are fewer distinctive events. For example, recalling details about what I did on my first day at work in my current job diminishes the longer I am in my current job because I have more memories of work days featuring similar activities from which to select which was my first day. Finally, often verbatim details are lost in the encoding process and inferences and interpretations are added and this can result in individuals ‘recalling’ in all sincerity events that did not actually occur. Such inferences may be added in response to the retrieval context (for example see Ref. [25]).

There are several processes involved in retrieval including: adopting a retrieval strategy, generating specific retrieval cues to trigger recall, remembering individual memories, and filling in partial memories through inference. Certain characteristics of the question and the material retrieved from memory can affect the completeness of the retrieval phase [26]. There are a number of reasons why respondents will not be able to recall an event, including:

- the item may never have reached long-term memory;
- the item may be irretrievable because the context of recall is different to the encoding context;
- the item may be difficult to distinguish from similar events or information;
- the item may be tainted with interference or contamination from another similar event (modified from Ref. [18]).

In designing questions it is important to assess how easily respondents will be able to retrieve the information required, to what level of detail, and to use context to help the respondent’s own recall strategies.

Judgement

The researcher, in designing a survey question, assumes that the respondent can provide the information being requested. Judgement forms an important part of the question-and-answer process because: the information being sought is often difficult to recall accurately (such as dates or frequencies); what can be recalled may be incomplete (such as recalling the details of a particular event); or, in the case of attitude or opinion questions, the question is asking the respondent to express a view or opinion on something which they may not have thought about (for some time) or in that context [22, 6].

Thus judgement can be seen as the process by which respondents formulate their answers to a survey question. This process involves respondents considering, for example, whether they understand the question, whether the question applies to their situation, whether it is asking for information they have, how detailed this information needs to be, how accurate it needs to be, whether they need to modify their answer to meet the perceived needs of the question and so on. These judgements may be made at any stage during the question-and-answer process, and can inform the comprehension, recall and response phases. For example, in being asked a question about how many times I have visited a doctor in the past 6 months, I may refine my interpretation of the terms ‘visited’ and ‘doctor’ in light of a first recall attempt. I may tighten or loosen my definitions depending on how easy it has been to retrieve the information required. I may consider that if I am having difficulty recalling the event then perhaps it happened infrequently. Alternatively I may be uncertain of the exact dates and thus make a judgement about whether I think the events I can recall happened within the reference period or not.

Judgmental heuristics

Cognitive ‘short cuts’ or heuristics are often employed by respondents when formulating answers to frequency questions. This is because, as noted above, memory is not perfect and decisions often have to be taken about how to compensate for incomplete or inaccurate information. For example, research has shown that the number of items to be recalled is the most important factor in

determining whether the respondent will adopt a counting or estimating strategy (for example, see Refs. [27, 28]). A number of different strategies for estimating answers to frequency questions have been identified and can be classified as follows:

- (1) recall of specific events, typically individual episodes;
- (2) estimation based on recall of summary information about the rate of occurrence of the event;
- (3) recall of an exact count or tally of events; and
- (4) estimation based on a general impression (modified from Tourangeau et al. [6]).

An alternative classification of judgemental heuristics is based on the work of Tversky and Kahneman [29, 30], which considers how available the information is to the respondent (ease of recall), how representative the information retrieved is of the 'normal' state of affairs and the use of context to anchor-and-adjust the answer. For example, in answering the question 'How many hours did you work last week?' respondents could opt for an easily available answer such as their contracted hours. Alternatively respondents may feel that their contractual hours are not representative of the true hours they work, and thus they may decide to average out the hours they worked over a number of weeks. Finally respondents may take their available or representative answer and adjust it in light of the survey context, the answer categories provided or the special circumstance of last week (see Sudman et al. [22] and Tourangeau et al. [6] for a review of Tversky and Kahneman's research on judgment).

Response

The final task described by the question-and-answer model is the response stage. There are two components involved in responding to the question: formatting and editing the response, and these two stages are described below.

Formatting response

Having formed a judgement the respondent then often has to fit his or her answer into one of the pre-specified answers being offered. This response formatting process is required where a closed answer is required, with the pre-defined answers

having already been designed by the researcher. For example, the following pre-specified answers could be offered for the question 'How often do you exercise?':

Everyday
 Every other day
 At least two or three times a week
 At least once a week
 Less often than once a week
 Never

The researcher's choice of response alternatives may affect the way the respondent decides to answer the question, and thus may affect the survey results [31]. Moreover the choice of response alternatives may also affect the way in which respondents interpret the question and the recall and judgement strategies they use. For example, the pre-specified answers provided for the question above about frequency with which people exercise could influence the respondent's answer by implying that it is common for people to exercise at least once a week. This is because only two of the six answer options provided refer to exercising less often than once a week, and one of those options is never exercising. The respondent may infer from the answer categories provided what the 'usual' frequency of exercise is for an average person, and thus they may use this inference to anchor and adjust their own answer to the question.

Editing the response

Finally respondents may want to edit their answers before they communicate it because they may want to conform to notions of social desirability and self-presentation. These effects may be more profound in face-to-face interviews than telephone or self-completion data collection methods.

The impact of social desirability factors on response is often limited to questions perceived by respondents as being sensitive and potentially threatening. However what constitutes a sensitive or threatening question, and thus what is socially desirable, will depend on the survey context – the mode of data collection, the characteristics of the interviewer, who else is present when answers are given apart from the interviewer, the content and purpose of the interview and so on (see Tourangeau et al. [6], for an overview of research in this area).

Applying theories of survey response – the cognitive methods toolkit

The value of theories which seek to explain the survey response process is in their application to the design and implementation of survey instruments and to reducing measurement errors. In particular they allow us to explore the individual components of the question-and-answer process: comprehension, retrieval, judgement and response.

Various cognitive methods have been developed and applied to the testing of survey instruments. These include cognitive interviewing, paraphrasing, card sorts, vignettes, confidence ratings and response latency timing [32–34]. The use of these methods, particularly cognitive interviewing, has become increasingly wide spread in the past 25 years with many western government statistical agencies now having cognitive testing facilities or laboratories [35]. The methods are summarised below.

Cognitive interviewing is a diagnostic tool for pre-testing survey instruments such as questionnaires. It focuses mainly on the questionnaire rather than the survey process, paying explicit attention to the mental processes respondents use to answer survey questions and thus allows covert as well as overt problems to be identified. It is qualitative and flexible in nature, being complementary to, rather than a replacement for, traditional field testing or piloting.

There are two main cognitive techniques:

- think aloud interviewing, and
- probing.

These two methods involve an interviewer asking the respondent about how she or he went about answering the survey question or completing a self-completion questionnaire. In the think-aloud approach the respondent is asked to ‘think-aloud’ as she or he answers the question or completes the questionnaire, whereas the probing method involves the interviewer asking specific questions or *probes* which are designed to elicit how the respondent went about answering the question. Whilst the think-aloud method is usually used concurrently, to collect information on what the respondent is thinking as she or he answers each survey question or completes the self-completion questionnaire, probing can be used either concurrently or retrospectively. The key differences between the think-aloud and probing techniques are:

Think-aloud	Probing
– Respondent-driven	– Interviewer-driven
– Lower burden on interviewer as respondent does most of the talking	– Lower burden on respondent, as respond to interviewer asks questions
– Can make the interview more difficult for the respondent	– Can make the interview easier for the respondent

The decision over whether to use the think-aloud or probing method is usually informed by the purpose of the test and what is being tested. Think-aloud tends to work better for self-completion questionnaires than for face-to-face interviews. This is because in a face-to-face interview it is harder for the respondent and interviewer to switch back and forth between the roles of survey respondent/cognitive respondent and survey interviewer/cognitive interviewer. However, both methods can be combined effectively.

Often standard probes are used to explore comprehension, retrieval, judgement and response processes [36]. Table 2 provides examples of the different types of probes used.

Paraphrasing involves respondents being asked to paraphrase the survey question, or rephrase it in their own words. This technique is princi-

Table 2. Examples of cognitive probes

Think-aloud/general	How did you go about answering that question? Tell me what you are thinking? I noticed you hesitated before you answered – what were you thinking about? How easy or difficult did you find this question to answer? Why do you say that?
Comprehension	What does the term <i>X</i> mean to you? What did you understand by <i>X</i> ?
Retrieval	How did you remember that? Did you have a particular time period in mind?
Confidence judgement	How did you calculate your answer? How well do you remember this? How sure of your answer are you?
Response	How did you feel about answering this question? Were you able to find your first answer to the question from the response option shown?

pally concerned with identifying comprehension problems and may be incorporated into a cognitive interview using such probes as: ‘Could you tell me in your own words what the question is asking you?’ or ‘How would you say that question yourself?’

Card sorts are used to explore how people group items together, and are useful in exploring and developing typologies and classifications. In a ‘free sort’ respondents group items that seem to naturally go together. In a dimensional sort respondents are asked to place items along a pre-specified dimension or dimensions.

Vignettes are short descriptions of hypothetical situations or scenarios, and are useful in understanding how respondents would answer questions about these situations and in showing whether the conceptual boundaries of the domain vary as between respondents.

Confidence ratings involve respondents being asked to rate the degree of confidence in their answers. As with paraphrasing, this technique can be incorporated into a cognitive interview using a probe such as: ‘How confident are you in your answer – where 10 is very confident and 0 is not at all confident?’

Response latency involves measuring the time elapsed between the presentation of the question and the indication of a response. The assumption here is that questions requiring more memory searching have longer response latencies.

Many of the cognitive methods described above were originally developed by psychologists to investigate how people solve problems or remember things. As mentioned earlier, these methods have been adopted by survey researchers to investigate sources of measurement error and to feed into the development of survey response theories. However the methods are still in the process of being refined to meet the needs of survey researchers, and whilst they have proved extremely useful tools for pre-testing survey instruments they are not without limitations.

Limitations of cognitive methods

Cognitive interviews are qualitative in nature, and thus whilst they can indicate the existence of a problem, they cannot provide quantitative infor-

mation on its extent or the size of its impact on survey estimates. To obtain this type of information alternative question-testing methods would be required. Furthermore, whilst cognitive methods can be used to evaluate existing forms of questions, and to test proposed revisions to the original questions, they cannot provide quantitative evidence on whether the revised version of the question is better than the original.

A further limitation is that the methods rely principally on respondents’ verbal reports of problems. Not all cognitive processes can be verbalised, as some happen so quickly they leave no trace in working memory [6]. For example, if you ask respondents their gender and then ask them how they went about answering that question they will not be able to tell you very much about the process: they just know what their gender is.

Critics of cognitive methods also cite that the methods can discriminate against less articulate respondents, who find it difficult to verbalise their thought processes and may also be less inclined to participate in cognitive testing experiments. Perhaps more worrying is the charge that the cognitive interviewing process can affect the way respondents answer questions (see Wilson et al. [37] for a summary of studies showing such an impact).

Finally, because cognitive methods are fairly new they are still relatively non-standardised [38]. As a result there are concerns that the results are not always reliable, particularly those based on cognitive interviews [39]. However the evidence in this regard is still patchy, principally because so few studies evaluating different pre-testing methods have been published to date.

Conclusions

Despite the reservations and limitations outlined above cognitive methods have greatly improved our understanding of sources of measurement error in quantitative surveys. They have also helped researchers develop better questions and survey instruments and are increasingly being used routinely to pre-test questionnaires. As such they should be seen as one of the components in the researcher’s tool bag for accessing whether the questionnaire does the task it is supposed to do.

References

1. Fowler FJ. *Improving Survey Questions: Design and Evaluation*. Applied Social Research Methods Series: 38. Thousand Oaks: Sage, 1995.
2. Fowler FJ, Mangione TW. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Applied Social Research Methods Series: 18. Newbury Park: Sage, 1990.
3. Oksenberg L, Cannell C, Kalton G. New strategies for pretesting survey questions. *J Off Stat* 1991; 7(3): 349–365.
4. Belson WA. *The Design and Understanding of Survey Questions*. Aldershot: Gower, 1981.
5. Clark HH, Schober MF. Asking questions and influencing answers. In: Tanur JM (ed.) *Questions About Questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation, 1992: 15–48.
6. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge: Cambridge University Press, 2000.
7. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl Cogn Psychol* 1991; 5: 213–236.
8. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opin Quart* 1987; 51: 201–219.
9. Cannell CF, Oksenberg L, Converse JM (eds). *Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971–1977*. Survey Research Centre, Institute of Social Research, University of Michigan. Ann Arbor, MI: University of Michigan, 1979.
10. Cantril H. *Gauging Public Opinion*. Princeton, NJ: Princeton University Press, 1944.
11. Payne SL. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press, 1951.
12. Rugg D. Experiments in wording questions: II. *Public Opin Quart*. 1941; 5: 91–92.
13. Sudman S, Bradburn NM. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine, 1974.
14. Jabine T, Straf M, Tanur J, Tourangeau R (eds). *Cognitive Aspects of Survey Methodology: Building a Bridge Between the Disciplines*. Washington, DC: National Academy Press, 1984.
15. Hippler HJ, Schwarz N, Sudman S (eds). *Social Information Processing and Survey Methodology*. New York: Springer-Verlag, 1987.
16. Tanur JM. Looking backwards and forwards at the CASM movement. In: Sirken MG, Herrmann DJ, Schechter S, Schwarz N, Tanur JM, Tourangeau R (eds), *Cognition and Survey Research*. New York: Wiley, 1999: 13–19.
17. Aborn M. CASM revisited. In: Sirken MG, Herrmann DJ, Schechter S, Schwarz N, Tanur JM, Tourangeau R (eds), *Cognition and Survey Research*. New York: Wiley, 1999: 21–38.
18. Tourangeau R. Cognitive sciences and survey methods. In: Jabine T, Straf M, Tanur J, Tourangeau R (eds). *Cognitive Aspects of Survey Methodology: Building a Bridge Between the Disciplines*. Washington, DC: National Academy Press, 1984: 73–100.
19. Esposito J, Jobe J. A General Model of the Survey Interaction Process. Bureau of the Census Seventh Annual Research Conference Proceedings. Washington DC: US Bureau of the Census, 1991: 537–560.
20. Schober MF. Making sense of questions: an interactional approach. In: Sirken, MG, Herrmann DJ, Schechter S, Schwarz N, Tanur JM, Tourangeau R (eds), *Cognition and Survey Research*. New York: Wiley, 1999: 77–93.
21. Suchman L, Jordan B. Validity and the collaborative construction of meaning in face-to-face surveys. In: Tanur JM (ed.) *Questions about Questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation, 1992: 241–267.
22. Sudman S, Bradburn NM, Schwarz N. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass, 1996.
23. Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. *Psychol Rev* 1973; 80: 352–373.
24. Anderson J. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press, 1983.
25. Loftus E, Palmer J. Reconstruction of automobile destruction: An example of the interaction between language and memory. *J Verb Learn Verb Behav* 1974; 13: 585–589.
26. Jobe JB, Tourangeau R, Smith AF. Contributions of survey research to the understanding of memory. *Appl Cogn Psychol* 1993; 7: 567–584.
27. Blair EA, Burton S. Cognitive processes used by survey respondents to answer behavioral frequency questions. *J Consum Res* 1987; 14: 280–288.
28. Sudman S, Schwarz N. Contributions of cognitive psychology to advertising research. *J Advertising Res* 1989; 29: 43–53.
29. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. *Cogn Psychol* 1973; 5: 207–232.
30. Tversky A, Kahneman D. Judgement under uncertainty: Heuristics and biases. *Science* 1974; 185: 1124–1131.
31. Schwarz N, Hippler HJ. Response alternatives: The impact of their choice and presentation order. In: Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S (eds), *Measurement Errors in Surveys*. New York: Wiley, 1991: 41–56.
32. Czaja, R. Questionnaire pretesting comes of age. *Market Bull* 1998; 9: 52–66.
33. Forsyth BH, Lessler JT. Cognitive laboratory methods: A taxonomy. In: Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S (eds) *Measurement Errors in Surveys*. New York: Wiley, 1991: 393–418.
34. Jobe JB, Mingay DJ. (1991) *Cognition and survey measurement: History and overview*. *Appl Cogn Psychol* 1991; 5: 175–192.
35. Schwarz N. Questionnaire design: The rocky road from concepts to answers. In: Lyberg L, Biemer P, Collins M, de Leeuw E, Dippo C, Schwarz N, Trewin D (eds), *Survey Measurement and Process Quality*. New York: Wiley, 1997: 29–45.

36. Willis GB. Cognitive Interviewing and Questionnaire Design: A training manual. Hyattsville, MD: National Centre for Health Statistics, 1994.
37. Wilson TD, LaFleur SJ, Anderson DA. The validity and consequences of verbal reports about attitudes. In: Schwarz N, Sudman S (eds), Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass, 1995: 91–114.
38. Willis GB, DeMaio T, Harris-Kojetin B. Is the bandwagon headed to the methodological promised land? Evaluating the validity of cognitive interviewing techniques. In: Sirken MG, Herrmann DJ, Schechter S, Schwarz N, Tanur JM, Tourangeau R (eds), Cognition and Survey Research. New York: Wiley, 1999: 133–153.
39. Presser S, Blair J. Survey pretesting: Do different methods produce different results? In: Marsden PV (ed), Sociological Methodology. Beverley Hills, CA: Sage, 1994; 24: 73–104.

Address for correspondence: Debbie Collins, National Centre for Social Research, 35 Northampton Square, London EC1V 0AX, UK
E-mail: d.collins@natcen.ac.uk