Paradigms of cognitive interviewing practice, and their implications for developing standards of best practice

Beatty, Paul

PARADIGMS OF COGNITIVE INTERVIEWING PRACTICE, AND THEIR IMPLICATIONS FOR DEVELOPING STANDARDS OF BEST PRACTICE

PAUL BEATTY

1. Introduction

Although cognitive interviewing has emerged as a major method to identify and correct problems with survey questionnaires, researchers who employ the method seem to lack consensus on several important points. There does not appear to be a commonly accepted definition of what cognitive interviewing is; we also lack detailed knowledge about what actually happens in cognitive interviews, other than the general notion that people "think out loud" and possibly probe for additional information about the meaning of responses, how people come up with answers, difficulties they have, and so on.

A definition that seems consistent with its most common application is as follows: cognitive interviewing entails administering a draft survey questionnaire while collecting additional verbal information *about* the survey responses, which is used to evaluate the quality of the response, or to help determine whether the question is generating the sort of information that its author intends. But beyond this general categorization, cognitive interviewing potentially includes different activities that may be based on different assumptions about the type of data that is being collected and the role of the interviewer in that process. For example, the verbal material generated by such interviews could consist of respondent elaborations regarding how they constructed their answers, explanations of what they interpret the questions to mean, reports of any difficulties they had answering, or anything else that sheds light on the broader circumstances that their answers were based upon. This material could be based on explicit follow-up questions (or probes) from an interviewer, or based on general instructions to "think out loud" as

much as possible. The interviewer herself could range from a relatively unskilled data collector to an expert investigator; the interview could be based on a scripted protocol, semi-scripted, or largely improvised based on the issues that emerge from discussion. Analysis may be based on systematic review of interview transcripts, or entirely from notes taken during the interview. The various permutations of these activities and their underlying assumptions could lead to quite different products, but all could apparently fall under the rubric of "cognitive interviewing."

Given such variety, it may be difficult to understand what someone means when claiming to have conducted cognitive interviews. Furthermore, a lack of consensus on objectives, procedures, or even general terminology can inhibit methodological developments. The major goal of this paper is to lay the groundwork for continued discussion about cognitive interviewing methodology by reviewing what it is, where it came from, and where it may be going. One potentially useful approach is to consider cognitive interviewing as falling into one of two major paradigms. The first of these paradigms has apparent roots in psychological laboratory methods, while the latter is more strongly rooted in the tradition of "intensive interviewing." After reviewing these paradigms, it will be possible to critically evaluate cognitive interviewing in its various guises – what it accomplishes, what it does not, and some of the key assumptions that its practitioners make when employing the method. The discussion will touch on numerous components of cognitive interview practice, including the role of the interviewer, selection of interviewees, data collection procedures, and evaluation of cognitive interview data.

2. Emergence of Cognitive Interviewing

By the mid 1980s, survey researchers had accumulated considerable knowledge about survey questions, but significant limitations remained – for example, they had limited knowledge about the mechanisms involved in response effects. Specific guidance for writing survey questions was still largely dictated by common sense and the experience of individual researchers. Basic field pretests appeared to be the most commonly used method for evaluating draft survey questions.

A seminar known as the first "CASM" meeting (for Cognitive Aspects of Survey Methodology) assembled survey researchers and cognitive psychologists in 1983 at St. Michaels, Maryland. The influential report from this meeting, *Cognitive Aspects of Survey Methodology: Building a Bridge Across Disciplines* (Jabine/Straf/Tanur/Tourangeau, 1984) proposed a number of important interdisciplinary collaborations. The report also introduced a four-stage model explaining how respondents are likely to answer survey questions: they must comprehend the question, recall relevant information, judge

the appropriateness of the information available to the particular question, and respond in the format provided (Tourangeau, 1984). This model, now widely adopted by methodologists (and expanded in Tourangeau, Rips and Raskinski, 2000), offers considerable help for both researchers and questionnaire designers, who can evaluate how well questions work with regard to each of these components. While these developments were taking place, researchers at the Center for Surveys, Methods, and Analysis (ZUMA) in Mannheim, West Germany, were considering similar ideas and held a conference of their own, with largely new participants. The resulting volume, *Social Information Processing and Survey Methodology* (Hippler/Schwarz/Sudman, 1987), presented some of the first substantive findings from collaborations between survey methodology and cognitive psychology.

Back in the U.S., the National Science Foundation (NSF) had funded several new collaborative research projects in the wake of the CASM meeting. One such project explored how principles of cognitive psychology might be applied in a laboratory setting toward the development and evaluation of questions on the National Health Interview Survey, which is conducted by the National Center for Health Statistics (NCHS) (Lessler/ Tourangeau/Salter, 1989). Additional funds from NSF were used to establish a "cognitive laboratory" at NCHS. Staff of this facility would evaluate and pretest questionnaires on a regular basis, in addition to investigating question-based response errors. Soon thereafter, similar laboratories were established at the Bureau of Labor Statistics and the Census Bureau, (Dippo/Norwood, 1992), to be followed by laboratories in academic and commercial research organizations (Forsyth/Lessler, 1991; Sirken/Schechter, 1999). Cognitive interviewing is the most common activity conducted in these laboratories.

3. Paradigms of Cognitive Interview Practice

It may be useful to distinguish between two general paradigms of cognitive interviewing. One involves a cognitive interviewer whose role is to encourage participants to verbalize their thoughts as they come to mind, but to intervene as little as possible in generating this verbal information. The other involves an interviewer who asks additional, direct questions about responses and who may assume greater responsibility for guiding a discussion about the basis for responses. The former paradigm relies almost entirely upon the think-aloud procedure, in which interviewers encourage participants to verbalize thoughts while answering questions (e.g., "tell me what you are thinking... how are you coming up with your answer to this?"). The latter paradigm relies more heavily on follow-up probes administered after the participant has answered the question (e.g., "can you tell me in your own words what that question was asking?"), although think-alouds may be encouraged to provide supplemental information. Below, I consider the origins and general parameters of both of these paradigms.

The "pure" think-aloud and non-intervening cognitive interviewer

The original paradigm of cognitive interviewing was explicitly psychological. Loftus (1984), elaborating upon ideas presented at the first "CASM" meeting, proposed that a technique known as *protocol analysis* could be adapted as a pretesting methodology for survey questions. The blueprint for this technique was developed by Ericsson/Simon (1980; expanded in 1993) and relies heavily upon *think-aloud* reports. Think-aloud reports were used to yield insights into the thought processes involved in participants' completion of certain tasks in a laboratory setting. Loftus reported that protocol analysis of think-alouds yielded information about how participants tended to retrieve memories of medical visits, and this information was used to develop question wordings that reflected these retrieval strategies. For example, she suggested defining the reference period of recall questions from a past date up to the present, rather than from the present backwards.

Early papers on cognitive laboratory methods (e.g., Royston/Bercini/Sirken/Mingay, 1986) suggest that initial cognitive interviews were based heavily, if not exclusively, upon thinking-aloud. In practice, this meant that cognitive laboratory participants would be asked to report what they were thinking while answering survey questions, and interviewers would simply remind respondents to continue providing such information as necessary. Think-aloud responses were therefore the unique data products of the interviews, and interviewer behavior was constrained accordingly.

An alternative paradigm: interviewers asking direct questions about responses

At some point, an alternative paradigm of cognitive interviewing emerged that expanded upon the use of "pure" think-alouds – in particular, allowing for the addition of direct probing by the interviewer. Apparently, the distinction between true think-aloud interviews and "intensive interviews" (which had been used to evaluate questionnaires prior to the CASM meetings – see DeMaio, 1983) became blurred, with both eventually falling under the header of "cognitive interviewing." It is easy to imagine how this could have occurred, especially if the intensive probes used could be construed as "cognitive" – addressing how terms were interpreted, how participants remembered certain facts, whether answers fit into available response categories, and so on.

This paradigm seems to have emerged gradually. Most descriptions of cognitive interviewing in the early 1990s (Forsyth/Lessler, 1991; Bercini, 1992) focus on think-alouds as the dominant component of cognitive interviewing, mentioning the possibility of probing for supplemental purposes – although Willis/Royston/Bercini (1991) suggest that both think-alouds and probing could both be viable alternatives. Later, Willis (1994)

proposed putting a greater emphasis on probing. Several other articles suggest that the trend toward acceptance of such activities continued: Gerber/Wellens (1997), for example, noted that cognitive interviewing had seemed to evolve from its original form to include "more probes and probes about meaning than was originally intended" (p. 35). Willis/ DeMaio/Harris-Kojetin (1999), noting that cognitive interviews are often called "think-aloud interviews," recommended that the latter term should be used more sparingly because think-aloud protocols were not necessarily the dominant component of cognitive interviews as currently practiced. And, Beatty (in press) concluded that cognitive interviews conducted in a study at NCHS relied heavily on interviewer adaptation to particulars that emerged in individual interviews.

Taken together, these works suggest that some practitioners of cognitive interviewing adopted a paradigm allowing for the collection of additional verbal material other than "pure" think-alouds, and at least in some cases, empowering the interviewer to guide the content of interviews. This is not to say that the use of think-alouds was abandoned, since virtually all descriptions of cognitive interviewing mention think-alouds as one possible component. It is also not to say that this paradigm completely replaced the alternative one firmly rooted in think-alouds, as Conrad/Blair/Tracy (2000), among others, clearly favor the original approach. Rather, another paradigm emerged, one that in practice owed relatively little allegiance to the procedures for verbal protocol analysis proposed by Ericsson and Simon.

The goal under both paradigms is to generate verbal information that is usually unseen in a survey interview in order to evaluate how well the questions are meeting their objectives. This puts both paradigms on an important common ground. Yet they are carried out differently and are based on some different assumptions, which may have implications regarding the data that they generate. It is therefore worthwhile to consider the rationales offered for both paradigms.

An assessment of the two paradigms

Advocates of the original paradigm propose that is has several advantages. One advantage proposed by Conrad/Blair/Tracy (2000) is that relying only on thinking-aloud avoids problems with artificiality that can arise due to interviewer probing. It is virtually indisputable that inserting probes into the middle of a survey interview alters the content and flow of the interaction. Such alteration of the realistic flow is the major reason why Oksenberg/Cannell/Kalton (1991) proposed that probing should be used after only a few questions per interview.

Forsyth/Lessler (1991) and van der Veer/Hak/Jansen (2002) are among those who propose an additional advantage: that think-aloud data are preferable because they are collected *during* the response process, and therefore have a certain purity that probe responses (provided *after* responding) do not. However, there is a considerable body of research beginning with Nisbett/Wilson (1977) that calls into question whether think-alouds are literal reflections of actual thought processes. More likely, they are re-constructions created after the fact. For the most part, they are likely to be reasonable reflections of actual processes (Wilson/LaFleur/Anderson, 1996) – but not necessarily literal representations. Furthermore, as Willis (in press) notes, Ericsson/Simon (1980) themselves did not insist upon exclusive use of thinking-aloud: their crucial point was that self-reported information should be in *short-term memory* (as opposed to long-term). From that perspective, reports based on probes immediately following questions are probably not much different than think-aloud reports. Both are approximations of "the real thing."

Advocates of the more probing-centered paradigm suggest that the alternative has particular advantages as well. One is that probing may provide necessary focus to cognitive interview interactions. Willis (1994) suggests that thinking aloud often leads participants to diverge onto irrelevant tangents. The most efficient way to correct this problem is through probes selected to re-focus attention on pertinent issues. Of course, doing so requires interviewer judgment. This is important, because it is not the use of probes per se that regain control of the interview, but an interviewer skilled at using the "right" probes. What Willis is really advocating is interviewer discretion in guiding the interview content.

Another potential advantage of probing is that it may have less of an impact on the response process than thinking-aloud. Although the actual content of probe responses is probably quite similar to that of think-alouds (see above), procedures for obtaining them are different; in the think-aloud case, participants at least attempt to provide some verbal information prior to responding to the question. According to Ericsson/Simon (1980), thinking-aloud should not interfere with the response process. However, Russo/Johnson/Stephens (1989) found that thinking aloud did have an impact on the accuracy of various mental computations; furthermore, Willis (1994) argues that thinking aloud increases the effort spent on creating a response, which has an unknown impact on the response process.

Perhaps the strongest justification for the more probe-based paradigm is that it generates verbal material that questionnaire designers find useful, but that may not emerge unless a cognitive interviewer specifically asks for it. As Willis (in press) observes, think-aloud

procedures were originally proposed by Loftus (1984) to shed light specifically on *retrieval* processes. However, the cognitive model proposed by Tourangeau (1984) also addresses comprehension, estimation/judgement strategies, and selection of particular responses. Even if think-alouds are effective at illuminating retrieval processes, they may not consistently generate information about these other issues. For example, participants might not explicitly consider the meanings of words or phrases while answering questions, and might not recognize their own misunderstandings. Also, Conrad/Blair/ Tracy (2000) note that think-alouds alone sometimes suggest a problem with a question without providing enough information to diagnose what the problem is. Probe responses might help to fill in this gap.

Today, the original distinction between the paradigms (thinking-aloud vs. probing) is probably not of central importance. Advocates of the original paradigm (Conrad/Blair/ Tracy, 2000) have conceded that probing makes important contributions, and advocates of the alternative (Beatty, in press) have acknowledged that probing can shape interview content in some undesirable ways. Most conceptualizations of cognitive interviewing today include some degree of probing. The primary distinction is now between an *unobtrusive* cognitive interviewer, who relies on standardized think-aloud protocols and possibly scripted probes, and an *active* cognitive interviewer, who is given more latitude to explore topics as they emerge within interviews. The practical decision has moved from whether or not to allow probes, to *how much* probing is appropriate, whether this probing should be standardized or determined by interviewer judgment (or to what extent), and how researchers should select the most appropriate probes for various purposes. The remainder of this paper considers how researchers might make these and other decisions in applying cognitive interviewing methodology.

4. Toward More Specific Practical Guildelines

While Forsyth/Lessler (1991), Willis (1994), and DeMaio/Rothgeb (1996), among others, have contributed significantly to establishing general parameters of cognitive interviewing, the literature is almost silent about many specifics regarding the design, implementation, and analysis of studies based upon this method. Guidance regarding the probing decisions mentioned above would be useful, as would guidance about the ideal background and training of interviewers, how many interviews are required to adequately test a questionnaire, and how participants should be selected.

Cognitive interviewers as data collectors vs. investigators

Tucker (1997) – in a position largely consistent with the original paradigm discussed earlier – calls for much greater standardization of cognitive interview procedures. Without this, he argues that "effective manipulation [of variables] will be impossible... the notion of falsifiability has no meaning... [and] the conditions necessary for generalizability will be absent" (p. 72). Conrad/Blair (1996) similarly argue that "rigorous experimental methods that rely on objective quantifiable data" are preferable to "largely impressionistic methods" that they suggest are generally used in cognitive laboratories (p. 8). Under this perspective, creative contributions from interviewers that lead to non-standardized behavior are undesirable. Given that interviewers would be constrained against improvising, the investigative burden is clearly on the front-end, meaning that researchers would need to determine in advance any issue they wished to probe about.

An alternative perspective is that interviewers themselves may serve as investigators. For example, Willis (1994) compares cognitive interviewers to "detectives" who rely at least partially upon improvisation in looking for clues about questionnaire problems. In subsequent work, he draws an analogy between cognitive and clinical interviews, which may be guided by intuition, experience, and flexibility (Willis, in press). This perspective forgoes consistency across interviews in favor of freedom to explore issues that emerge in discussions with participants. Presumably its major advantage is that it allows interviewers to explore issues that emerge within the interview, were not anticipated in advance, and might be missed through more scripted interviews. While this perspective does not preclude identifying some issues to be watchful for in advance, it does place considerable trust in the interviewer's ability to notice potential problems and to conduct *emergent probing* in ways that shed light on the sources of these problems. Thus the interviewer takes on some of the role of investigator as well as data collector.

These two perspectives might call for very different sorts of cognitive interviewers, with potentially different skills, backgrounds, and training. The skills necessary for data-collection cognitive interviewers might not be much different than those of survey interviewers – e.g., they would require training in general procedures, but not in the subject matter being investigated (Fowler/Mangione, 1990). They would not have to know why think-alouds or probes were being administered – only to recognize when participants were providing adequate think-aloud or probe responses.

For investigative cognitive interviewers, such skills would be necessary but not sufficient, since they would at least partially determine the content of the interview. In doing so, such interviewers might need to draw upon knowledge of the objectives of the questions, potential types of cognitive or communicative errors that could affect the accuracy of

survey responses, and familiarity with various options for eliciting useful verbal material from participants. Strangely, the literature on cognitive interviewing does not seem to address the appropriate background of such cognitive interviewers. A solid grounding in survey methodology would probably be useful, since interviewers would generally be working in an environment geared toward producing survey statistics, and presenting their conclusions to professionals in that field. It would be useful for advocates of this type of cognitive interviewing to think clearly about what other skills or background are most desirable for identifying effective interviewers.

What to ask: the selection of probes

As discussed earlier, most recent conceptualizations of cognitive interviewing involve probing to some degree. If the interviewer is also an investigator, then she may select some of these probes herself; if a data collector, then the probes may be selected for her. But either way, someone must choose what probes are used. Although cognitive interviewing literature provides many examples of possible probes, it provides little guidance regarding which probes are likely to be most effective for various purposes. A few basic guidelines are available: for example, Willis (1994) notes that probes should not suggest a "correct" answer, a principle that also applies to survey questions. Foddy (1998) concludes that specific probes such as "what does [term] mean to you?" are more effective than general ones such as "what were you thinking when you first answered the question?" In another recent study, Beatty (2002) found that participants answered probes about the meaning of terms differently when they were administered alone than they did within the context of a particular survey question. However, these sort of recommendations appear to be uncommon, and almost completely missing from published literature in this area.

Cognitive interviewers may be able to obtain some guidance about how to choose "good" probes from literature on qualitative interviewing, which may include lessons on what to ask, how to ask it, and how to make sense of narrative data. For example, Weiss (1994) suggests that interviewers generate narrative by asking about specific events rather than generalized experience. Holstein and Gubrium (1995) encourage interviewers to be on the lookout for "confusion, contradictions, ambiguity and reluctance" as signs that "meanings are being examined, reconstituted, or resisted" (p. 79). In the case of cognitive interviews, such instances might call for additional probing. Variants of qualitative interviewing are also employed by anthropologists, and some guidance may be obtainable from that field as well. For example, Gerber (1999) notes that anthropologists might explore whether terms are "culturally inappropriate" for a particular population. But rather than simply asking a participant what a term such as *self-reliance* means, an anthropologist might

explore its meaning in different contexts, e.g., with regard to child rearing, older family members, or welfare recipients. This might suggest that general cognitive interview probes such as "what does this term mean to you?" might be less effective than specific ones exploring how a term is used in a participant's life.

Who to interview, and how much interviewing to do

Cognitive interviewing literature pays even less attention to issues of how to select samples of participants and how to determine when an adequate number of interviews have been completed. The lack of attention to this issue is likely a consequence of its association with psychological laboratory methods, which have often placed little emphasis on such matters. Cognitive interview practitioners generally acknowledge that participants are chosen by convenience and that such samples are "not designed to be representative [of any larger population], but to reflect the detailed thoughts and problems of the few respondents who participate in [cognitive interviews]" (DeMaio/Mathiowetz/Rothgeb/Beach/Durant, 1993).

Other than that, the only specific guidance that seems to be available is that some demographic variety of respondents is desirable, and that participants should include people relevant to the topic of the questionnaire being tested (Willis, 1994). One clear consequence of such sampling is that cognitive interviewing can never determine the extent of questionnaire problems in a population. Still, some sampling considerations could help to strengthen claims that a reasonably thorough effort has been made to identify the most pressing problems with a questionnaire.

For example, participants could be selected to cover as much of a questionnaire's conceptual terrain as possible. If questionnaires include skip patterns that lead to various branches, the sample should be sufficiently diverse to explore all of these different paths. Whatever topic the questions focus on (e.g., health insurance), the sample should cover a variety of circumstances relevant to that topic (e.g., people with a variety of health insurance situations). Within those parameters, it also seems desirable to select participants representing some demographic variety. Practitioners should not operate under the illusion that such diversity ensures "representativeness"; it only casts a wider net over varying circumstances, maximizing the chances that discovery will be effective. Similarly, interviewing in multiple locations could improve the variety of circumstances that are captured in testing.

As for what constitutes an adequate cognitive interview sample size, little guidance has been offered on this point either – often, literature in this area simply acknowledges that samples are small. Several researchers report that cognitive interviews are commonly

divided among several “rounds” consisting of about 10 interviews each (Willis, 1994; McColl, 2001). Such small rounds of interviews are considered sufficient to identify some questionnaire problems, at which point questionnaires can be revised and tested again in subsequent rounds. This iterative approach seems useful, but it still leaves open the question of whether researchers can determine when they have conducted enough rounds of interviews to stop the process. Some qualitative researchers make decisions regarding when enough interviews have been conducted based on the idea of *category saturation* (Strauss/Corbin, 1990). Put simply, this means that the researcher identifies groups of people most relevant to the study and conducts interviews with members of each until they yield relatively few new insights. In other words, operating under a principle of *diminishing returns* may be effective. Operating in this manner makes a very important assumption: that the most critical questionnaire problems will be revealed quickly from virtually any group of relevant participants. This assumption is most likely to hold up if participants reflect a range of experiences that a question attempts to measure, and also represent at least an attempt to obtain some demographic diversity. These are not guarantees of representativeness. Rather, they are guidelines that maximize the chances of discovering potential questionnaire problems as efficiently as possible. Still, greater attention to how participants are selected and how many of them should be interviewed could maximize cognitive interviewing’s potential to quickly home in on the most significant problems with a particular questionnaire.

Evaluating evidence from cognitive interviews

Whether cognitive interviews are conducted based on a fairly standardized protocol or with greater interviewer flexibility, the result is still verbal text that needs to be evaluated to determine whether or not a question poses a problem for respondents. One advantage of fairly standardized protocols is that they allow for more systematic analysis. For example, Conrad/Blair (1996) propose that verbal protocols be coded in a table with “types of problems” on one axis (lexical, temporal, logical, etc.), and “response stage” (understanding, task performance, and response formatting) on the other. Whenever problems were observed answering a question, they would be coded in the appropriate category. Of course, the success of this procedure (and others like it) is based on the assumption that the interviewing technique brings cognitive problems to the surface so that they can be observed. It also assumes that an analyst will be able to make enough sense of these verbalizations to code them appropriately.

As we have seen, some may counter-argue that additional probing from a skilled investigator brings enough additional material to the surface to justify the lack of standardization – that is, the above procedure would miss some important observations.

The resulting data may be harder to interpret and it may be harder to judge whether a problem is "real." One possibility for evaluating questionnaire problems is that attempts should be made to link them to characteristics of the question. For example, consider the survey question "Thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?" Schechter/Beatty/Willis (1998) concluded that respondents might have a difficult time answering this. The conclusion can be backed up by the following process of reasoning:

- *observation of a problem*: several participants could not provide codeable responses (a number between zero and thirty), even when probed.
- *studying the specifics of the problem*: some participants indicated that the question did not allow them to answer in meaningful terms ("a day is part good and part bad – you can't characterize it as one or the other"); others complained about response task difficulty ("I don't do bookkeeping on this"), especially given complicated health status.
- *identifying the question characteristic that is the source of the problem*: the question is based on the assumption that a "day" is a reasonable metric, but it may not be for people with varying-quality days.
- *determining whether this is this generalizable*: it seems reasonable that this problem could recur, e.g., for people who experience health problems that bother them at various times during the day, or people with multiple intermittent health problems.

Claiming that this process found "proof" of the problem would be overstating the evidence. However, a reasonable case could be made that the problem is likely to be found in respondents with similar circumstances, and is created by a faulty assumption about the way individuals think about their health. Note also that the evidence is not linked to the *number* of participants who report a particular problem. Whether it takes many or a few participants to construct such an argument, it needs to be evaluated based on logical merits. It is conceivable that a solid argument about a questionnaire problem could be constructed around a single case, or that such an argument might fail to materialize around several.

Error in cognitive interview analysis

Still, since the method does rely heavily upon human judgment, the possibility that cognitive interviewing could lead to conclusions that are incomplete, misleading, or incorrect must be addressed. There are several possibilities for error: cognitive interviews could identify problems that would not turn out to be "real" in surveys; cognitive interviews could fail to identify problems that exist in actual survey administration; and, cognitive interview findings might be inconsistent when conducted by independent

groups of researchers. The first two could be considered problems with the validity of the method and might be respectively classified as *errors of commission* and *errors of omission*. The third could be considered to be a problem with the reliability of the method.

Practitioners of cognitive interviewing make several assumptions to defend themselves against the possibility of errors of commission. One is that *cognitive interviewing finds problems that will carry over to actual surveys.* Unfortunately, there is often no obvious way to verify that hypothesized problems are "real." Logical arguments may have to do, and researchers will have to determine for themselves whether they find such arguments to be meritorious. However, there has been at least one attempt to verify that cognitive interview findings were borne out by field data (Willis/Schechter, 1997). They revised questions based on cognitive interview findings, and the results were more plausible than statistics from the original question – however, such efforts are expensive and the results are not always conclusive.

Another potential concern is that cognitive interviewing could *fail* to find problems that would actually appear in survey interviews. Actually, there is no reasonable way that cognitive interview practitioners could claim to have found all problems with a questionnaire. Its usefulness is based on the assumption that the most egregious problems will become evident in most groups of participants who are reasonably appropriate to the topic of the survey. Interviewing often concludes based on a subjective judgment that interviews are yielding diminishing returns. However, there is always the possibility that one additional interview could yield a significant new insight, or that an additional interviewer would be more likely to notice additional problems. By the same token, claims that a questionnaire has "no problems" are impossible – the strongest claim that could be made is that no problems have (yet) been discovered.

Finally, there is always the possibility that independent groups of cognitive interview practitioners might not reach the same conclusions about a questionnaire. However, it would probably not be unusual for different groups to discover different insights, especially if interviewers were operating under an "investigator" paradigm. Differences could be a function of different interviewer backgrounds and sensitivities to various sorts of problems. Hopefully, however, the findings of different groups would not be wildly incompatible (e.g., with one group finding a series of comprehension and recall problems, and another finding no problems even after many interviews).

On a related note, one potential view of cognitive interviewing is that its objective is to identify and eliminate all of the problems with a question, and that by doing so we can find the "ideal" question to ask. However, an alternative view is that researchers could

reach very different conclusions about the quality of a certain question, both of which could be correct. Rather than attempting to find the "right" way to ask a survey question, cognitive interviewing may be more suited to helping researchers assess the advantages and disadvantages of asking questions in a certain manner – and helping researchers make decisions based on the potential errors they are more comfortable with. Questionnaire designers actually make these decisions all the time (e.g., adding verbiage to a question might clarify its intent, but doing so could also make it longer and more burdensome). It may be that cognitive interviewing proves to be useful simply because it provides information to make such design decisions as logically as possible, some of which might not be available through other pretesting methods. In other words, cognitive interviewing may be less suited to finding the "best" questions than guiding "best informed" design decisions.

It seems unlikely that cognitive interviewing will generate "reliable" findings in the sense that survey researchers might use the term (i.e., each set of interviews identifies the same set of problems with questions). When findings are different, yet not necessarily contradictory, this may indicate that no one set of findings is complete – the different findings should be examined to see whether they complement or refute each other. Findings that are difficult to reconcile might indicate either faulty reasoning by analysts, or that interviewing has not yet yielded an adequate understanding of responses associated with a question. The former case calls for a closer look at the data, while the latter indicates a need for continued data collection.

5. Conclusions and overall assessment

Since its inception in the mid 1980s, cognitive interviewing has become a prominent method for survey questionnaire development and evaluation. It should be clear from the preceding review that cognitive interviewing is not so much one clearly defined method as a loose collection of several potential activities. These activities have a great deal in common – all involve the collection of verbal material beyond a simple survey response, which is used to evaluate whether questions are capturing information as intended. Yet there are also important differences in both philosophy and practice.

Reasonable arguments have been offered for both the data-collector and investigator paradigms of cognitive interviewing. Clearly both positions offer some advantages. Both also have shortcomings that could be addressed more thoroughly by their respective advocates. Discussions about best practices need to continue. With such continued discussions, researchers should be better equipped to use cognitive interviewing to help create questions that are clear, pose memory and recall tasks that respondents can

reasonably be expected to accomplish, and that allow respondents to express their answers accurately.

References

Beatty, P., 2002: Cognitive Interview Evaluation of the Blood Donor History Screening Questionnaire. In Final Report of the AABB Task Force to Redesign the Blood Donor Screening Questionnaire. Report submitted to the U.S. Food and Drug Administration.

Beatty, P., (In press): The Dynamics of Cognitive Interviewing. In S. Presser/J. Rothgeb/ M. Couper/J. Lessler/E. Martin/J. Martin/E. Singer (eds.), Questionnaire Evaluation and Testing Methods Development. New York: John Wiley and Sons.

Bercini, D.H., 1992: Pretesting Questionnaires in the Laboratory: An Alternative Approach. Journal of Exposure Analysis and Environmental Epidemiology, 2, 241-248.

Conrad, F./Blair, J., 1996: From Impressions to Data: Increasing the Objectivity of Cognitive Interviews. 1996 Proceedings of the Section on Survey Research Methods, Volume 1, pp. 1-9. Alexandria, VA: American Statistical Association.

Conrad, F./Blair, J./Tracy, E., 2000: Verbal Reports Are Data! A Theoretical Approach to Cognitive Interviews. In Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference. Washington, D.C.: Office of Management and Budget.

DeMaio, T.J. et al., 1983: Approaches to Developing Questionnaires. Statistical Policy Working Paper 10. Washington, D.C.: Statistical Policy Office, U.S. Office of Management and Budget.

DeMaio, T.J./Mathiowetz, N./Rothgeb, J./Beach, M.E./Durant, S., 1993: Protocol for Pretesting Demographic Surveys at the Census Bureau. Washington, D.C.: U.S. Bureau of the Census.

DeMaio, T.J./Rothgeb, J., 1996: Cognitive Interviewing Techniques: In the Lab and in the Field. In N. Schwarz/S. Sudman (eds.), Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass.

Dippo, C.S./Norwood, J.L., 1992: A Review of Research at the Bureau of Labor Statistics. In J. Tanur (ed.), Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage.

Ericsson, K.A./Simon, H.A., 1980: Verbal Reports as Data. Psychological Review, 87, 215-251.

Ericsson, K.A./Simon, H.A., 1993: Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press.

Foddy, W., 1998: An Empirical Evaluation of In-Depth Probes Used to Pretest Survey Questions. Sociological Methods and Research, 27, 103-133.

Forsyth, B.H./Lessler, J.T., 1991: Cognitive Laboratory Methods: A Taxonomy. In P. Biemer/R.M. Groves/L. Lyberg/N. Mathiowetz/S. Sudman (eds.), Measurement Error in Surveys. New York: Wiley.

Fowler, F.J./Mangione, T.W., 1990: Standardized Survey Interviewing: Minimizing Interviewer-Related Error. Newbury Park, CA: Sage.

Gerber, E.R., 1999: The View from Anthropology: Ethnography and the Cognitive Interview. In M. Sirken/D. Herrmann/S. Schechter/N. Schwarz/J. Tanur/R. Tourangeau (eds.), Cognition and Survey Research. New York: Wiley.

Gerber, E.R./Wellens, T.R., 1997: Perspectives on Pretesting: 'Cognition' in the Cognitive Interview? Bulletin de Methodologie Sociologique, 11, 18-39.

Hippler, H./Schwarz, N./Sudman, S. (eds.), 1987: Social Information Processing and Survey Methodology. New York: Springer-Verlag.

Holstein, J.A./Gubrium, J.F., 1995: The Active Interview. Thousand Oaks, CA: Sage.

Hyman, H.H./Associates, 1975 [1954]: Interviewing in Social Research. Chicago: University of Chicago Press.

Jabine, T.B./Straf, M.L./Tanur, J.M./Tourangeau, R. (eds.), 1984: Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, DC: National Academy Press.

Lessler, J.T./Tourangeau, R./Salter, W.: Questionnaire Design in the Cognitive Research Laboratory. Vital and Health Statistics, Series 6, No. 1. Hyattsville, MD: National Center for Health Statistics.

Loftus, E.F., 1984: Protocol Analysis of Responses to Survey Recall Questions. In T. Jabine/M. Straf/J. Tanur/R. Tourangeau (eds.), Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, DC: National Academy Press.

McColl, E., 2001: Protocol for Cognitive Testing of Global Health Status Rating Items. Unpublished manuscript.

Nisbett, R.E./Wilson, T.D., 1977: Telling More Than We Can Know: Verbal Reports on Mental Processes. Psychological Review, 84, 231-259.

Oksenberg, L./Cannell, C.F./Kalton, G., 1991: New Strategies for Pretesting Survey Questions. Journal of Official Statistics, 7, 349-365.

Royston, P.N./Bercini, D./Sirken, M./Mingay, D., 1986: Questionnaire Design Research Laboratory. 1986 Proceedings of the Section on Survey Research Methods, pp. 703-707. Alexandria, VA: American Statistical Association.

Russo, J.,/Johnson, E.,/Stephens, D., 1989: The Validity of Verbal Protocols. Memory and Cognition 17: 759-769.

Schechter, S./Beatty, P./Willis, G.B., 1998: Asking Survey Respondents About Health Status: Judgment and Response Issues. In N. Schwarz/D. Park/B. Knäuper/S. Sudman (eds.), Cognition, Aging, and Self-Reports. Philadelphia, PA: Psychology Press.

Sirken, M./Schechter, S., 1999: Interdisciplinary Survey Methods Research. In M. Sirken/D. Herrmann/S. Schechter/N. Schwarz/J. Tanur/R. Tourangeau (eds.), Cognition and Survey Research. New York: Wiley.

Strauss, A./Corbin, J., 1990: Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park, CA: Sage.

Tourangeau, R., 1984: Cognitive Sciences and Survey Methods. In T. Jabine/M. Straf/J. Tanur/R. Tourangeau (eds.), Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, DC: National Academy Press.

Tourangeau, R./Rips, L./Rasinski, K., 2000 : The Psychology of Survey Response. Cambridge: Cambridge University Press.

Tucker, C., 1997: Methodological Issues Surrounding the Application of Cognitive Psychology in Survey Research. Bulletin de Methodologie Sociologique, 11, 67-92.

van der Veer, K./Hak, T./Jansen, H., 2002: The Three-Step Test Interview (TSTI): An Observational Instrument for Pre-testing Self-Completion Questionnaires. Paper presented at the Questionnaire Development, Evaluation, and Testing Conference in Charleston, South Carolina, November 2002.

Weiss, R.S., 1994: Learning from Strangers: The Art and Method of Qualitative Interviewing Studies. New York: The Free Press.

Willis, G., 1994: Cognitive Interviewing and Questionnaire Design: A Training Manual. Cognitive Methods Staff Working Paper Series, No. 7. Hyattsville, MD: National Center for Health Statistics.

Willis, G. (in press): Cognitive Interviewing Revisited: A Useful Technique, in Theory? In S. Presser/J. Rothgeb/M. Couper/J. Lessler/E. Martin/J. Martin/E. Singer (eds.), Questionnaire Development Evaluation and Testing Methods. New York: John Wiley and Sons.

Willis, G./DeMaio, T./Harris-Kojetin, B., 1999: Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques. In M. Sirken/D. Herrmann/S. Schechter/N. Schwarz/J. Tanur/R. Tourangeau (eds.), Cognition and Survey Research. New York: Wiley.

Willis, G.B./Royston, P.N./Bercini, D., 1991: The Use of Verbal Report Methods in the Applied Cognitive Laboratory. Applied Cognitive Psychology, 5, 251-267.

Willis, G./Schechter, S., 1997: Evaluation of Cognitive Interviewing Techniques: Do the Results Generalize to the Field? Bulletin de Methodologie Sociologique, 11, 40-66.

Wilson, T./LaFleur, S./Anderson, D., 1996: The Validity and Consequences of Verbal Reports About Attitudes. In N. Schwarz/S. Sudman (eds.), Answering Questions: Methodology for Determining Cognitive Processes in Survey Research. San Francisco, CA: Jossey-Bass.

Contact

Paul Beatty
National Center for Health Statistics
3311 Toledo Road
U.S.A. - Hyatsville, MD 20782
U.S.A.
email: pbb5@cdc.gov