Cognitive Interviews: Best Practices

Institute for Social Research/ Survey Research Center

Cognitive Interviews: Best Practices
DATE: September, 2007

Cognitive Interviews: Best Practices[1]

Background

Spurred by a desire to develop SRC's capability to conduct cognitive interviews, a questionnaire pre-testing method, the *SRC Technical Infrastructure Group* (TIG) provided funding to F. Conrad and others in the *Survey Methodology Program* in collaboration with members of the *Survey Research Operations Unit*, to conduct a cognitive interview study to test and document interviewer effects in cognitive interviews. The study took place during the summer of 2006.

The current document describes best practices for conducting cognitive interviews, based on the lessons learned from the study funded by TIG and experiences from other recent cognitive interview studies conducted at SRC and by F. Conrad and colleagues at several institutions. The best practices take the form of "frequently asked questions," each followed by recommendations or guidelines for practice.

As SRC continues to conduct carefully designed studies of cognitive interviewing, the hope is that continued lessons learned will be integrated into this living document.

[1] Source: Project funded by the SRC Technical Infrastructure Group to support work by Conrad, F., Blair, J., and Tourangeau, R. in collaboration with the SRC Survey Research Operations (K. Alcser). Best Practices lead authors: Kirsten Alcser and Frederick Conrad with input from Johnny Blair, Rachel Levenstein, and Roger Tourangeau (May 2007).

Cognitive Interviews: Best Practices
DATE: September, 2007

1. *What Are Cognitive Interviews*?

Cognitive interviewing is a questionnaire pretesting technique in which respondents think aloud while answering draft questions and interviewers' probes about their thinking. The technique usually exposes problems with respondents' comprehension of questions but can also expose problems in how they construct answers and choose response options. "Cognitive interviewing techniques are used to study the manner in which targeted audiences understand, mentally process, and respond to the materials we present – with a special emphasis on potential breakdowns in this process" (Willis, 2005, p. 3). "The cognitive interview is usually conducted within a broader testing environment, as part of a sequence of activities that, in aggregate, are referred to as cognitive testing" (Ibid, p. 5).

2. *When Should Cognitive Interviews Be Conducted*?

Cognitive interviews are most likely to help when administered during questionnaire development, i.e. after questions have been drafted and reviewed by experts in both questionnaire design and the subject matter but prior to field pretesting. They are not the occasion to explore how respondents think about the survey *concepts* as is the case for focus groups, but are well suited to evaluating how respondents interpret and answer specific questionnaire items.

3. *What Kinds of Problems Are Cognitive Interviews Best at Identifying*?

In several studies (e.g. Presser & Blair, 1994; Conrad & Blair, 2004), the most frequent types of problems uncovered by cognitive interviews concerned the meanings of individual words (comprehension) and the way they are connected (logic). Comprehension problems might be evident when a respondent explicitly indicates he or she does not know what a word means or the respondents' think-aloud reveals they have misinterpreted a word. Logical problems may be produced when words are connected by several conjunctions (*and*s and *or*s).

4. *How Many Interviews*?

It is common in practice to conduct only a small number of cognitive interviews. For example, Willis, 2005 (p.7), indicates that between 5 and 15 interviews are typically conducted per interviewing round. However, Blair, Conrad, Ackermann & Claxton (2006) have observed that the number of unique problems identified in cognitive interviews increases with the sample size. In their study, the number of problems identified continued to increase through samples of size 50 and these included serious problems as rated by expert judges. Their data suggest that the rate at which additional interviews uncover new problems begins to slow down at around 25 or 30 interviews. So if one has limited resources, we believe they will be most effectively spent on about 25 - 30 interviews; a larger number of interviews will uncover more problems (including serious ones) but at an

increased cost per problem. While more studies with a variety of questionnaire material certainly need to be done before we can really understand the relationship between sample size and problem detection, as of now, this is our best recommendation. As with any other form of data collection, increasing quality in cognitive interviewing (e.g. by increasing sample size) may increase costs. We believe that the standard practice of using very small samples may reduce cost but sometimes reduce confidence in conclusions.

Once the problems identified in the set of cognitive interviews have been documented, the researcher can revise the original questionnaire so that respondents no longer experience these problems. Of course, revising the questionnaire does not guarantee the problem has been fixed; worse, the repair process may introduce new problems. The only way to evaluate the consequences of the repair effort is to conduct another round of interviews similar in size to the initial round. Willis (2005) indicates that iterative testing with cognitive interviews is a common practice and, while we know of no empirical evaluations of its effectiveness, it seems on intuitive grounds much more powerful than a single round of testing. Thus we recommend at least two rounds (with 25-30 respondents each), but more if at all possible.

Within a round of cognitive interviews it also makes sense to stop and take stock after the first few interviews have been conducted. This could be quite informal, e.g., debriefing the interviewers, or more formal, e.g., coding recorded verbal reports. This can make it evident, for example, that the interviewer probes are eliciting useful information and, if not, provides an early opportunity to revise the procedure.

5. Who Are the Interviewers?

In many organizations that conduct cognitive interviews, the interviewers tend to have advanced degrees in one of the social sciences (Blair and Presser, 1993). We believe this not essential. In two recent studies (Blair, et al., 2006; Levenstein et al., 2007), the cognitive interviewers were skilled production interviewers who were trained for these studies to administer the survey questions and a set of scripted probes verbatim. (We will discuss the pros and cons of scripted probes below; but for now we note that these seemed effective at detecting problems.) This suggests to us that, in general, i.e. beyond methodological experiments, production interviewers can be trained to use this technique. This approach (production interviewers administering scripted probes) may lead to higher quality data than the more typical approach in which professional social scientists improvise probes. There is some evidence that the more typical approach produces verbal reports from respondents that are not very reliably interpreted by independent judges (Conrad and Blair, 2004).

In contrast, in the recent studies which employed production interviewers administering scripted probes (Blair et al., 2006; Levenstein et al., 2007),

agreement between judges was quite high. Furthermore, in the second study, the researchers tried to vary the apparent expertise of the interviewers but it appeared to have no effect. In an introductory description, the same interviewers were sometimes described as highly experienced, sometimes not. In addition, when the interviewers were described as more experienced, they also wore white lab coats to appear more clinical, and conducted the interviews in front of technical looking equipment. Yet this variation in apparent expertise had no impact on the effectiveness of the cognitive interviews in detecting problems, further evidence that production interviewers, i.e. without professional social science credentials, can be quite appropriate.

6. *How Scripted Should the Interviews Be*?

We believe that production interviewers are skilled at delivering questions and keeping respondents on task, and so can also administer scripted probes effectively; but they are less experienced with improvising in the interview settings and so probably are not as effective at crafting probes on the fly. This in part is due to the subject matter expertise and survey methodology background required to recognize that respondents are having trouble and determining the best way to expose the problem. We have not experimented with giving greater discretion to production interviewers so cannot rule out their use of improvised probes; but our current experience suggests this approach may not fit their background and skill set.

The optimal procedure for identifying successful cognitive interviewers from a production interviewer pool is not well established and there is no relevant data of which we are aware. Success may depend on factors such as listening ability, tenure (or, perhaps, lack of tenure as those who have worked fewer years in a strictly standardized, production interviewing environment may be better able to adapt to the demands of collecting think alouds and attending to the respondents' thinking), and experience conducting face-to-face interviews (cognitive interviews are typically conducted face-to-face).

We think it is possible to involve cognitive interviewing staff without background in survey methodology (e.g., production interviewers) in developing the probes. The interviewers would review the questionnaire and suggest probes to explore problems they believe might be present. Then, together, the interviewers and supervisors would settle on a set of probes to be administered by all of the interviewers. In general, when using scripted probes, we recommend reassessing the script midway through the project so that as additional problems become apparent in the interviews, probes for these problems can be added; conversely, probes that are not informative can be removed from the script.

Interviewers with more research background, especially in survey methodology, should be able to formulate some probes on the spot to supplement the scripted probes. Our work has suggested these probes should be (1) tied to explicit

evidence of problems in respondents' think alouds, i.e. interviewers should not probe about problems for which there is no prior evidence (Conrad & Blair, 2004), and (2) interviewers should ask generically about response problems without specifying a problem (Conrad et al., 2006). For example, if a respondent pauses for a long time before answering or says he or she doesn't know how to answer, we would recommend the interviewer inquire why the respondent paused or did not know how to answer. What we would NOT want an interviewer to do is to propose and seek confirmation of a problem (e.g. "Did you have trouble answering because none of the response options fit your true feelings?").

7. *What Should the Interviewer Document*?

The primary job of the cognitive interviewer is to keep the interview flowing and encourage the respondent to think out loud. An additional task that seems potentially helpful, but which we have not yet tested, is to ask the interviewers to make some basic observations about the interaction. This might be structured so that after the exchange surrounding a question the interviewer provides some simple, yes-no data in response to statements such as "The respondent had a problem with the question." "The respondent asked me to re-read the question." "I had trouble reading the question." This kind of data could help point out the exchanges that particularly warrant a closer listen by someone analyzing the recorded interviews after the fact. Moreover, the task could help keep the interviewer focused and, by requiring him or her to write something, she will appear to be more engaged than if she simply listens. However, because this places extra demand on the interviewer it may not be practical for novice cognitive interviewers. It might be helpful, particularly for more experienced cognitive interviewers, to provide a small number of problem categories that they can check when there is appropriate evidence. This might help the analyst in quickly identifying certain problems.

8. *Who Are the Respondents*?

It is our experience that most people are able to think aloud, though some may require more practice and encouragement than others, so most anyone can be a cognitive interview respondent. If there are eligibility requirements for the production survey, e.g. more than 40 years of age or low income, these should be mirrored in the recruitment procedure for the cognitive interviews. Then, if the researchers have any hypotheses about likely problems which suggest that some groups should be more or differently affected than others, this should be tested by recruiting some respondents from within the groups suspected to experience the problem and some who are not group members. In the absence of such group-specific hypotheses, there are no particular requirements on who is recruited.

9. *How Are Respondents Recruited*?

There are many methods for recruiting participants for cognitive interview studies, but the appropriate method really depends on the particular survey. One reason for this is that, if the survey being planned concerns a special population, the cognitive interview participants should resemble that population. For example, if the study targets urban poor, or elderly, then it makes less sense to recruit from on-line resources like Craigs List than by posting fliers in community centers or making cold calls in certain neighborhoods. Among the possible sources we have used and have had positive experiences with are: Craigs List, fliers, email samples, newspaper advertisements, U-M on-line subject pools like the Psychology department's subject pool, the Turner Geriatric Center, UM-Engage, etc. Obviously, screening can refine the pool from any one source to better match the study goals.

10. *How Should the Results Be Analyzed?*

Cognitive interviews produce largely verbal and, therefore, qualitative data. We propose that data of this sort are best analyzed by classifying, i.e. coding, all of the verbal reports into categories of problems and then describing how the verbal reports fit into those categories. This can be time consuming – in our experience about 2 to 3 times the duration of each interview – but we believe it is worthwhile because it makes it possible check reliability and summarize results. If a relatively large number of interviews are conducted, then it is appropriate to report frequencies and possibly percent of occurrence for each category. If the number of interviews is smaller, then it is more appropriate to describe the types of problems found and any clear patterns among them. The particular set of categories will depend on the goals of the study but the main requirement is that they allow the verbal data to be objectively summarized. The categories may be reusable across studies (e.g. semantic versus recall problems) or specific to the content of questions (e.g. for questions about income from multiple sources respondents may have problems conducting mental arithmetic).

In order to give credibility to the coded results, some measure of inter-coder agreement or reliability, e.g. Cohen's Kappa, is required. Because coding is a time consuming process, computing reliability for a subset of verbal reports, e.g. all items from only some interviews, can reduce the resources required. We have found that cognitive interviews tend to produce lower agreement scores than one would like but are higher (though still not high) when there are some restrictions on interviewer behavior such as scripting the probes (Conrad & Blair, 2004; Levenstein et al., 2007). In light of our recommendation for scripting probes, we suggest that a criterion for agreement is kappa scores in the range of .61 and higher (see Everitt & Haye, 1992 for more details). Whatever the criterion, some reassurance that the coding results are general across coders is a prerequisite for

subsequent analyses. There is some evidence from open responses in production interviews that the longer the units of text being coded the lower the reliability (e.g. Conrad & Couper, 2001). Because the verbal reports in cognitive interviews are often quite long, it might increase coding reliability if they are decomposed into smaller units.

11. Implications for Practice

While we are reluctant to prescribe exactly how any one cognitive interviewing study should be conducted, we can recommend several guidelines at this point including:

1. Do at least 25 cognitive interviews; if extensive changes are made to the questionnaire, do a second round of testing with at least 25 more.

2. It is okay to use production interviewers rather than specially trained "cognitive interviewers".

3. Use scripted, general probes like "How did you come up with that answer?" or "Did you have any problems answering that question?" rather than probes about specific problems as this may promote "false alarms." When the respondents give specific signs of difficulty, have interviewers improvise general probes ("Why did you pause so long before giving your answer?") but otherwise limit the interviewers' discretion in how and when to probe.

4. Recruit participants who are similar to the target population.

5. Assess inter-coder reliability at the early stages of analysis (e.g. categorization) of verbal reports, intervene if necessary to increase reliability, and assess reliability at the end of the process.

REFERENCES

Blair, J., Conrad, F., Ackerman, A. C., & Claxton, G. (2006). The effect of sample size on cognitive interview findings. *Proceedings of the American Statistical Association, Section on Survey Research Methods.* Alexandria, VA: American Statistical Association.

Blair, J., & Presser, S. (1993). Survey procedures for conducting cognitive interviews to pretest questionnaires: A review of theory of practice. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Conrad, F. G., & Blair, J. (2004). Data quality in cognitive interviews: The case of verbal reports. Ch. 4 in Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (eds.) *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.

Conrad, F. G. & Couper, M. P. (2001). Classifying open ended reports: Coding occupation in the current occupation survey. *Proceedings of the Federal Committee on Statistical Methodology Research Conference* Friday, All Sessions, pp 21-30.

Everitt, B. S., & Haye, D. F. (1992). *Talking about Statistics: A Psychologist's Guide to Data Analysis*. New York: Halsted Press.

Levenstein, R., Conrad, F., Blair, J., Tourangeau, R., & Maitland, A. (2007). "The effect of interviewer behaviors in cognitive interviews: A signal detection approach." Paper presented at the meeting of the American Association for Public Opinion Research, Anaheim, CA.

Presser, S., & Blair, J. (1994). Survey pretesting: "Do different methods produce different results?". *Sociological Methodology*, *24*, 73-104

Willis, G. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* Thousand Oaks, CA: Sage Publications.

ADDITIONAL RESOURCES

Beatty, P. C., & Willis, G. (2007). "Research Synthesis: The Practice of Cognitive Interviewing". *Public Opinion Quarterly*, 71, 287-311.