

STANDARDIZED COGNITIVE TESTING: WILL QUANTITATIVE RESULTS PROVIDE QUALITATIVE ANSWERS

Carol Cosenza, Center for Survey Research
Carol Cosenza, CSR, University of Massachusetts Boston, 100 Morrissey Blvd, Boston, MA 02125

Key Words: Cognitive Testing, Question Evaluation

Using cognitive testing as part of the pre-field evaluation of a survey allows researchers to learn how the cognitive tasks posed by the questions are being handled. We can learn whether respondents understand our questions, how they figure out what information they need, and how they come up with their answers. (Fowler, 1995; Forsyth & Lessler, 1991) One of the most challenging aspects of cognitive interviewing is trying to make sense of all the information gathered during the interview.

BACKGROUND

There are many methods to collect data from cognitive interviews. Depending on who does the interview, researchers may get not only different points of views, but perhaps even different findings. Cognitive psychologists, project staff members, question evaluation specialists, and specially-trained interviewers can all conduct cognitive interviews. Each of these types of people brings special skills to the table, and each has their own costs. (DeMaio & Rothgeb, 1996; Cosenza & Fowler, 2000)

Methods of collecting data from cognitive interviews:

Just as there are benefits and drawback for each type of cognitive interviewer, there are benefits and drawbacks to the different ways researchers attempt to gather information collected during the interviews. (Willis, DeMaio, & Harris-Kojetin, 1999; Torangeau, Rips, & Rasinski, 2000) Sometimes individual summaries are written about each interview, highlighting what the interviewer feels is important. This allows the researcher to read an overview of the interview, which can then be looked at more in-depth if desired. Often, interviews are audio- or video-taped. Although this is helpful to check out certain issues or concerns, sometimes the researcher does not have the time or inclination to go through hours and hours of tape. If the tapes are transcribed, the problem simply changes from hours of tape to reams of paper through which to wade.

Interviewer debriefings are also very common. These debriefings usually consist of systematically going through the test questions and sharing observations and experiences that interviewers had when asking the questions and probing respondents about them. This gives the researcher a chance to hear about all the interviews at once, allows questions to be asked of the interviewers as a group, and also allows interviewers to hear how the other interviews

went. Issues that an individual interviewer may have thought to be idiosyncratic and might not have mentioned may be found to be more common when the group talks.

Most researchers use some combination of these methods, hoping to get the best parts of each without all of the drawbacks. Still, since cognitive interviews are usually not standardized, there can be no single method that provides systematic information retrieval from these unsystematic interviews.

ZUMA The Center for Survey Research and Methodology (ZUMA) in Mannheim, Germany has been working with a three-stage standardized question evaluation process. Part one is an expert review of the questions, followed by a standardized evaluation interview, and finally a field pretest. The standardized evaluation interview is a modified cognitive interview. When administered by interviewers (instead of senior staff members) the evaluation interview consists of the questions to be tested and pre-written standardized follow-up questions. Interviewers ask all questions exactly as written and do not engage in any spontaneous probing. Researchers learn about the cognitive issues from the quantitative results of the all the questions in the evaluation instrument.

Center for Survey Research, University of Massachusetts Boston At CSR, our usual cognitive interviewing protocol consists of a semi-structured cognitive instrument. It contains the test questions, the objective of the cognitive test, and possible probes an interviewer might use (or begin with) to get to that objective. Hoping to build on the work from ZUMA, we decided to compare a cognitive instrument that relies entirely on standardized follow-up questions that are as closed-ended as possible with our usual protocol where interviewers are free to innovate and cognitive probes require more narrative responses.

RESEARCH DESIGN

The questions being tested were part of a survey instrument to measure people's attitudes and beliefs about the benefits and risks of screening tests for cancer. Seven interviewers completed 21 interviews in March and April of 2001.

Three senior interviewers, specially trained in cognitive testing techniques, conducted 10 cognitive interviews using our usual protocol. These interviews were conducted in-person and were audio-taped. Interviewers attended a briefing that explained the purpose of the study and the specific cognitive concerns for each test question. The interviews had a semi-structured protocol that allowed interviewers leeway in what questions to ask and when to ask them. The interview schedule included the questions to be tested and a series of follow-up probe questions. Interviewers were also permitted, even encouraged, to ask additional probes to clarify the way that respondents understood the questions or the meaning of their answers. After completing their interviews, interviewers participated in a group debriefing session with project staff. This interviewer debriefing constituted the primary source of information about the questions and any question problems.

The test condition consisted of four members of CSR's professional interviewing staff who completed 11 interviews by telephone. These interviews were also audio-taped. The test protocol consisted of an alternate form of the cognitive interview that consisted solely of standardized cognitive probes. The key feature was that the cognitive follow-up questions were entirely pre-scripted and were to be read exactly as worded - just like any other question in a standardized interview. The answers would be given in a structured format. Thus, the answers themselves would provide the "results" of the cognitive interviews. The telephone interviewers were not debriefed; information from these interviews came from the tallies of answers to the structured cognitive probes.

RESEARCH QUESTION

Our research question was to see if the results of the standardized telephone interviews, conducted by interviewers with no special cognitive training who would not be debriefed, lead the researchers to the same conclusions about how the cognitive tasks are performed as the more traditional cognitive interviewing techniques.

SELECTED FINDINGS

EXAMPLE 1:

TEST QUESTION: How much do you worry about getting heart disease?

COGNITIVE GOAL: Is "heart disease" consistently understood?

What we learned in the debriefing: Respondents included everything from "clogged arteries," to "high cholesterol," to "leaky valves," and "heart attacks".

Standardized Probe: Would you say "worrying about heart disease" is the same as or different than "worrying about a heart attack?"

answered "SAME" - 5 respondents

answered "DIFFERENT" - 6 respondents

In this example, pre-identifying the problem - concern that respondents would hear "heart disease" and think about "heart attack" - allowed us to write a specific question that addressed our concern. The fact that almost half of the respondents answered the standardized probe as "same" and that the interviewers in the debriefing also mentioned it as a problem, lead us to realize that the question was not consistently understood.

EXAMPLE 2:

TEST QUESTION: Let's say that 100 women your age are diagnosed and treated for breast cancer. On average, how many of those women, do you think, will die from breast cancer in 5 years? Would you say more than 50, 25-50, 10-25, or fewer than 10?

COGNITIVE GOAL: How does the Respondent figure out her answer?

What we learned in the debriefing:

Respondents seemed confused and really didn't know how they chose their answer. Some talked about having read magazine articles or hearing about it somewhere. However, the interviewers felt that overall people were basically guessing.

Standardized Probe:

On a scale of 0 to 5 where 0 means totally unsure and 5 means totally sure, how sure are you of your answers to these questions?

average answer: 3.3

who didn't answer "5": 8 respondents (out of 11)

From both protocols we learned that respondents did not have the knowledge to answer this question. Although all of the respondents answered the original test question, when probed about their choice, interviewers felt that most were just guessing. The quantitative results from the standardized probes highlighted a problems and at the same time took the subjective issues off the table. We did not have to rely on interviewer instincts and observations such as the respondents “seemed confused” or “hesitated before answering” as possible signs of a problematic question.

EXAMPLE 3

TEST QUESTION: Again, thinking about your most recent mammogram, did you make a decision about whether to have that mammogram?

COGNITIVE GOAL: Did the Respondent really make a decision about having her last mammogram?

What we learned in the debriefing: Respondents said they made a decision, but after discussing it, we found that most simply had the mammogram their doctor suggested. Many actually had automatic appointments set up for yearly mammograms. The interviewers felt that for most it was not a real “choice.”

Standardized Probe: Do you think the decision was about when to have a mammogram or whether to have one at all?

answered “WHEN TO HAVE IT” - 8 respondents
answered “WHETHER TO HAVE IT”- 1 respondent

In this situation, the standardized probe gave us additional information that was not gathered by the unstructured interview. We found that even though the question asked about “whether” to have the mammogram, respondents were hearing - and answering - about “when”. The responses they were giving did not match the objective of the question. Throughout the interview, we learned that asking about “whether to have a mammogram” is a concept that is very hard for women to comprehend. The answer was so ingrained that it seemed too obvious a question to ask. However, asking about “when” or “how often” to get mammograms seem to be more manageable for most respondents.

EXAMPLE 4

TEST QUESTION: Some doctors feel that women older than 70 may not need to have mammograms every year. How likely is it that (_____) is a reason a doctor may not recommend mammograms - very likely, somewhat likely, a little likely, not at all likely?

- concern about the cost of mammograms
- scientific studies show that testing women over 70 does not decrease the risk of a woman dying of breast cancer
- concern about radiation exposure
- concern about the number of false alarms or false positives in women over 70

COGNITIVE GOAL: How does the Respondent understand this question?

What we learned in the debriefing: Respondents had a very difficult time with this question. They were confused and really didn’t know what we were asking about. Interviewers also talked about some respondents’ reactions to what they felt was implicit ageism in the question. Some felt that the question was an example of not valuing older women and were offended by it.

Standardized Probe: On a scale of 0 to 5 where 0 means the question is totally confusing and 5 means the question is totally clear, how would you rate this question?

average answer: 3.9
who didn’t answer “5”: 5 respondents (out of 11)

While we could take the results of the standardized probe as a flag that there was a problem with understanding the question, we could not pick up the unanticipated issue of how the respondents felt about the question in general. This unintentional finding from the flexible interview helped us better understand what one of the problems with the question was, not just that the question had a problem.

CONCLUSIONS

1. **For pre-identified concerns, we learned the same things from both the narrative and the structured interviews.** For concerns about concrete issues, such as definitions, comprehension, and recall issues, a structured cognitive probe seems to work just as well, if not better, than the open-ended responses.

2. **There are many benefits of using structured closed-ended probes.** A standardized interview is easier to administer. While a cognitive psychologist or senior project staff member may be good on their feet and figure out how to tailor each interview to the particular respondent, time and money often prevents these types of people from doing many cognitive interviews. By standardizing the instrument, general field or phone interviewers could ask these questions and more interviews could be done in a shorter amount of time, thus providing more data to help evaluate the questions at less cost. The standardization also makes it easier to compare results across interviews. When using typical cognitive interviewing techniques, or even our semi-structured probes, the results are still in a narrative form. Debriefings as a main source of gathering cognitive interview results are a much more complex and less systematic approach.

3. **Unstandardized cognitive interviews give us insights into the respondent and how he or she handled the question and answer process of the interview.** By listening to interviewer observations, researchers can learn more about how the process went. While a quantitative response, like a tally from behavior coding, can alert the researcher to a potential problem, it is in talking to the interviewers that unexpected results are found. By allowing exploratory questions and non-standardized probing, the researcher is not bound only to learn about those issues that were pre-identified. Unanticipated issues, concerns, and problems can then be considered in the evaluation process. Flexible interviews provide an analytical richness missed in a pre-written protocol.

4. **It is very hard for respondents, or interviewers, to judge how cognitively difficult a question is.** All researchers know of questions that are poorly written or cognitively complex but are still answered by respondents. Several times in our standardized interview we asked respondents to rate how confusing certain questions were or how sure they were of their answers. When we looked at the number of respondents who did not rate the question as “totally clear” or that they were not “totally sure,” we could see that some questions were truly more difficult than others. However, overall, people were very unlikely to say they didn’t understand something. After the rating questions, we asked why the question was confusing. In general, we found that respondents were not very good judges of this. While they might say there was a problem, it was a very difficult task for them to explain their own comprehension problems. The most common response was “I didn’t understand it.”

Our specially-trained interviewers also have a hard time figuring out the “why” part of the comprehension problem. They often know that respondents had a difficult time with a question. However, they are not very skilled

at figuring out why there is a problem or how it might be fixed.

As more and more studies routinely use cognitive testing, research in the field of question evaluation must continue to evolve. Who does the interviews, how many are done, in what order the cognitive questions get asked, and how structured the protocol is, can all influence the information learned through this process. Research, comparing different types of cognitive testing and in what situation these different types work best, needs to continue.

There is an enormous amount of information that is gathered during each cognitive interview. How that information gets summarized, and eventually used, is a very important step in the process that we should not overlook.

REFERENCES

- Cosenza, C. and Fowler, F.J. (2000). Prospective questions and other issues in cognitive testing. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 994-997.
- Fowler, F.J. (1995). *Improving survey questions: design and evaluation*. Thousand Oaks, CA: Sage Publications.
- DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz and S. Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, pp. 177-195. San Francisco: Jossey-Bass.
- Forsyth, B. H., and Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*, pp. 15-36. San Francisco: Jossey-Bass.
- Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.
- Willis, G. B., DeMaio, T., and Harris-Kojetin, B. (1999). Is the bandwagon headed to the methodological promised land? Evaluating the validity of interviewing techniques. In Sirken, M. B., et al (Eds.), *Cognition and Survey Research*, pp. 133-154. New York: Wiley.

ACKNOWLEDGMENT: I would like to like to gratefully acknowledge Peter Preuffer and Margrit Rexroth at ZUMA (Mannheim, Germany). A draft version of a paper in preparation, “Two-Stage Pretesting,” served as an inspiration and model for this research and paper.