COGNITIVE ISSUES AND METHODOLOGICAL IMPLICATIONS IN THE DEVELOPMENT AND TESTING OF A TRAFFIC SAFETY QUESTIONNAIRE

Susan Schechter and Paul Beatty, National Center for Health Statistics
Alan Block, National Highway Traffic Safety Administration
Susan Schechter, 6525 Belcrest Road, #915, Hyattsville, MD 20783

KEY WORDS: Cognitive research; traffic safety

Motor vehicle crashes continue to be one of this country's most serious safety problems. In 1993, there were more than 40,000 deaths and 3 million people injured due to motor vehicle crashes, resulting in tremendous personal costs and billions of dollars in health care and lost productivity (NHTSA, in press). Obviously, efforts to improve this situation could have far reaching benefits, both in our quality of life and in our nation's economic health.

The mission of the National Highway Traffic Safety Administration (NHTSA) is to help reduce deaths, injuries, and other unintended personal and economic costs of our highway transportation system. Agency success in achieving this goal depends heavily on the quality of data upon which it bases its programmatic decisions. In particular, NHTSA requires data that identifies barriers to public acceptance of safety behaviors. One way NHTSA will obtain this information is by conducting a national telephone survey to collect data on public knowledge, attitudes, and behavior regarding key occupant protection and other traffic safety issues. A national probability sample will be drawn and the 20 minute survey will be administered using computer-assisted telephone interviewing.

Instrument

For this survey, NHTSA requested that the National Center for Health Statistics (NCHS) collaborate in all aspects of questionnaire design and development--from refining survey objectives and drafting the initial set of questions to testing and finalizing the instrument in the NCHS Questionnaire Design Research Laboratory. It was clear from the outset that this new instrument would be challenging to develop on several accounts. First, there were a variety of survey objectives across many different topics. Questions providing measures of behavior, attitudes, and knowledge were to comprise the instrument framework. NHTSA also wanted to measure risk and other safety perceptions and was hoping to look at causality. For example, NHTSA not only wanted to know how often people wore seat belts, what they knew about seat belt laws, and their opinions about the laws, but the Agency also wanted to learn the reasons behind a person's decision to wear or not wear a seat belt and the factors that might influence this decision making. Added complexity to the instrument was also anticipated as some measures would likely be sensitive to context and social desirability effects. Safety-related behaviors and attitudes were clearly the central theme of the instrument. Thus, respondents might be reluctant to, for instance, acknowledge unsafe behavior or state opposition to safety measures such as seat belt laws.

Three-stage approach

In developing the approach for this project, various cognitive research methods used in questionnaire design were considered (Tourangeau, 1984; Lessler, Tourangeau, and Salter, 1989; Forsyth and Lessler, 1991; Willis, Royston, and Bercini, 1991). A three-stage research methodology was developed to meet the challenges anticipated with the project. For the first stage, questionnaire designers worked with subject matter experts to complete comprehensive cycles of expert analyses on draft versions of questions. Second, face-to-face cognitive interviews were conducted in the NCHS Questionnaire Design Research Laboratory to identify conceptual problems and to examine cognitive difficulties with questions and response alternatives. Third, simulated survey interviews by telephone, followed by face-to-face retrospective interviews, were also conducted in the laboratory prior to finalizing the questionnaire. The following is a discussion of the rationale, method, and results of each of these stages.

Stage 1: Expert analysis

During stage 1, a cognitive expert analysis of versions of draft questions was conducted. An expert analysis of a questionnaire is distinguished from the more common "technical review" by several factors. Based on the authors' experiences, the purpose of a technical review is to obtain comments regarding any major problems with either the wording of questions or the general formatting of the instrument. It is not a well-defined process and is not governed by rigid standards or rules.

In a cognitive expert analysis, the procedures are not well-defined either. A critical point of differentiation, however, between the technical review and the expert analysis is that the latter focuses on detecting potential response error through the application of principles of

cognitive psychology. Each question is analyzed from the respondent's perspective in an effort to understand the response tasks being imposed (Forsyth and Lessler, 1991). The analysis is comprehensive in nature and looks beyond the wording of particular questions to examine, for example, appropriateness of question structure and response alternatives, logical ordering of information, correct usage of skip patterns and so forth. While not a new technique, the authors' perception is that expert analysis is under-utilized due to the time-consuming nature of iterative review cycles and the need to have analysts with experience in cognitive aspects of survey measurement.

To illustrate the value of expert analysis, it may be helpful to examine the process of questionnaire design outside the influence of cognitive science. Based on the author's observations, typically the primary questionnaire designer is often times a subject matter expert who develops the first series of questions. The designer then sends the draft instrument to other subject matter specialists and perhaps to survey researchers for review and comment. Each subject matter specialist reviews those questions pertinent to his or her area of expertise for content and accuracy of terms and phrases. The survey researchers focus their review more on the instrument itself, looking for inconsistencies across questions, logical ordering of questions, awkward wording, and accuracy of skip patterns. Comments are sent back to the questionnaire designer, who incorporates the comments and distributes a revised version for perhaps one more cycle of review and comment. The instrument is then ready for a field pretest.

Conducting expert analyses requires that the questionnaire designer draw heavily from the principles of cognitive psychology in order to view the instrument from the respondent's perspective and simulate the response process without benefit of an interview (Forsyth and Lessler, 1991). In this first stage of questionnaire design, the primary focus is to identify cognitive problems with draft questions. Approaching the analysis from the respondent's point of view, each question is analyzed with the following in mind:

- How will a respondent interpret the meaning of a question?
- What sorts of retrieval tasks are being imposed?
- What kinds of response strategies will a respondent be likely to use?
- What process will a respondent likely go through in deciding on the right answer to the question?

Revisions are then made to the questions, new versions of questions are sent to subject matter experts and other researchers for further review and comment, and additional refinements to the instrument are made. These cycles, or rounds, continue until all the evident questionnaire design problems are rectified.

For this project, conducting cycles of expert analyses allowed for the identification of basic cognitive issues that did not need to be revealed later in an interview in order to be discovered and corrected. For example, a critical survey measure was whether or not the public knows they should wear a seat belt even when driving in a car having an air bag. The subject matter specialist recommended the question, *"If a vehicle has an air bag, does the law require a driver to also wear a seat belt?"* In analyzing the question, we realized that it asked the respondent to recall knowledge about seat belt laws and how the laws apply to vehicles with air bags. In actuality, the sponsor was not interested in measuring knowledge of the law, but rather, knowledge about whether seat belts should be worn in vehicles with air bags. To meet the survey objective, the proposed revision was, *"True or false...If a car has a driver side air bag, I don't need to wear my seat belt when driving."*

Another example of questionnaire improvement during expert analysis was revision to an item seeking to measure the degree of confidence respondents have in emergency medical services (EMS) personnel. The question proposed by the subject matter specialist was, *"Regardless of the type of medical emergency, are you confident that the EMS personnel would know what to do?"* A number of problems were identified when analyzing this question. First, it did not ask the respondent to recall a particular experience but rather required accessing a global opinion about one's confidence in EMS personnel. But, because the response scale did not allow for incremental measures of confidence, the respondent would be forced to choose between a yes answer (total confidence) or a no answer (no confidence). Also, even though the interviewer would define EMS for the respondent earlier in the interview, the term was still considered too technical to use. The proposed revision to this question was *"Regardless of the type of medical emergency, how confident are you that the ambulance or other emergency workers would know what to do...very confident, somewhat confident, or not very confident?"*

The examples above only serve to illustrate the sorts of improvements made to an instrument during an expert analysis stage. In addition to refining the wording, length, and logical ordering of questions, other types of problems that might have resulted from

poor comprehension, difficulty in recall and the use of inappropriate response strategies and heuristics can be identified and corrected. These problems would likely have been detected in other pretests (whether laboratory or field). However, we suggest that it is more efficient and effective to identify and resolve these sorts of problems as early in the questionnaire development process as possible.

Stage 2: Face-to-face cognitive interviews

For the second stage of questionnaire design, three rounds of face-to-face cognitive interviews were conducted in the questionnaire design laboratory with a total of 28 subjects. Subjects were recruited from an advertisement placed in a local newspaper as well as from flyers posted in neighborhood stores and social service agencies. In order to ensure adequate testing for a series of child safety seat questions, subjects who responded to the advertisement or flyers were screened so that at least half the subjects scheduled for a round of interviews had children under the age of 6. A targeted recruitment of lower socioeconomic subjects was also used in later rounds. The interviews lasted one hour and subjects were paid $30.

Using techniques common to questionnaire design laboratories, subjects were encouraged to think aloud while answering, and intensive probing techniques were used in order to discover cognitive issues (e.g., Tourangeau, 1984; Lessler, Salter, and Tourangeau, 1989; Willis, Royston, and Bercini, 1991). Results from these rounds of testing resulted in numerous changes to the instrument.

Even though driving is a common behavior, a lack of shared terminology in a number of topical areas was observed. For example, subjects had difficulty specifying the type of seat belt they wore when driving. This was a critical screening question as seat belt usage may be dependent on the type of seat belt one has. A terminology deficiency in describing a device people use or see nearly every day simply was not anticipated. When answering the question, *"What kind of seat belt is in the front seat of the car you usually drive?"* virtually every subject initially responded with an awkward hand motion to non-verbally indicate where their seat belt crossed their body. Verbatim responses included, *"I don't know what you call it; the thing that goes across here automatically; it's the one that you pull above the door."* These responses indicate that people think of their seat belt as just that--a seat belt--and they do not discriminate further. In revising this question, specific questions were suggested to establish whether the belt was a lap belt only, a shoulder belt only, or both, and then to identify whether it was automatic, manual or both.

A second critical observation during the cognitive interviews was the identification of errors and inconsistencies in reports of everyday behaviors. For example, it was expected that when subjects were asked frequency of seat belt or car seat use, they would access a general representation of usage and not respond based on a calculation of all recalled seat belt usages. However, it was not expected that the overwhelming majority of subjects who confidently said they wore their seat belt *"all the time"* would easily remember a time in the recent past that they did not wear it. Of even greater interest was that most of these subjects did not seem to think they had given an inconsistent or contradictory answer. In probing, subjects articulated that their interpretation of the response category *"all the time"* was that it represented the rule, not the exception, and that an *"all the time"* response should not be taken literally. To them, their answer was accurate and they did not think the exceptions to the rule needed to be considered. Rather than modify the response option to *"almost always,"* the sponsor added the probe *"When was the last time you did not wear your seat belt?"* into the survey instrument.

Stage 3: Telephone interviews with face-to-face debriefings

Questionnaire design concluded with two rounds of laboratory telephone pretests supplemented with face-to-face retrospective probing. Fifteen subjects were recruited for this stage of testing. Subjects came to the laboratory, met the interviewer, were asked to sit in a private room, and were called on the telephone by the interviewer from a different room. This testing stage called for the interviewers to administer the instrument without interruption or cognitive probing, so that the flow and timing of survey administration could be evaluated. Following the telephone interview, the subjects joined the interviewer and discussed the basis for answers to specific questions through face-to-face probing.

One of the more interesting observations during this stage was the propensity some subjects had to recall knowledge when asked to report an attitude. The following question and associated responses illustrates this observation: *"I'm going to read two types of seat belt laws and I'd like you to tell me which you favor. The first allows police to stop a vehicle solely for observing that adults violate seat belt laws. The second requires that the vehicle first be stopped for some other violation before ticketing for seat belt violations. Which do you favor?"* Subjects had considerable difficulty providing opinions about seat belt laws, even when

specific information about the laws was provided. The question assumes that respondents know details about laws and have existing attitudes about them. On the telephone, most laboratory subjects responded with phrases such as, *"The first one"* or *"The last one you said."* These responses cued the interviewers to problems with the question, as few subjects actually repeated the kind of law they favored. During the debriefing, the question was repeated for subjects and they were asked to paraphrase the question. Only a few subjects were able to do this correctly, which is even more significant given that it was the second time they had heard the question. Subjects were also probed to elaborate on their opinions of the different laws. The majority of subjects discussed their knowledge of seat belt laws rather than their preference for one law over another. Responses included, *"Yes, I think police are allowed to stop you if they see you without a seat belt on..."* and *"Oh, police give tickets if they don't see you wearing a belt."* Interviewers had to specifically instruct the subjects to think about which of the two alternative laws they preferred rather than what they thought reflected their local law. Once instructed, subjects were able to express some opinions, but most acknowledged guessing or a high degree of uncertainty.

Discussion

Selecting the right research methods

For this project, a multi-stage approach led to an effective and efficient design of the questionnaire. By extending the traditional technical review into a cognitive expert analysis, each draft version was analyzed to not only identify the more typical questionnaire problems related to structure and logic, but to detect cognitive problems that would hamper comprehension, recall, and response strategies. Working closely with the sponsor, many potential response issues were resolved prior to laboratory testing.

It should be noted that even though for purposes of this project the expert analysis was identified as the first stage of questionnaire design, expert analyses can and often times should continue as the questionnaire advances in other design stages. That is, expert analysis does not need to be limited to only the initial stage of design. Particularly when an instrument is being developed for the first time, rather than a revision of a prior data collection instrument, cognitive expert analyses should be a part of each stage throughout the questionnaire design and testing process.

Cognitive interviewing, as expected, revealed significant cognitive and structural problems with the instrument. Revisions were tested in successive rounds and the instrument continued to be refined throughout this process. The final stage of telephone interviewing and retrospective probing in face-to-face debriefings allowed us to simulate administering the instrument and to study cognitive flaws one more time.

In order to determine the kinds of cognitive research methods most appropriate for a given questionnaire design project, future research should focus on techniques to classify specific cognitive issues which surface in different stages of design. For example, are different classes of cognitive problems uncovered depending on the type of pretesting method used? Clearly, an examination of that question leads one to look at the other cognitive research methods that were available but not chosen for this project. For example, what criteria should survey researchers apply when trying to decide when to use a particular method, such as focus groups or behavior coding? Further research into the particular advantages and disadvantages of each technique will help practitioners to better conduct questionnaire pretesting.

Minimizing recall problems

Results of this project suggest several considerations for designing questionnaires to minimize recall problems. For example, Jobe, Tourangeau, and Smith (1993) discuss the use of instructions which clearly define the intent of the question and can reduce reporting errors. This would increase the length of survey administration time, but the benefits of longer questions may outweigh the disadvantages of possible response error.

Also, further research should be conducted on the wording and ordering of response alternatives to determine memory search impact. In the example of subjects who said they wore their seat belt all the time and then recalled a recent time during which they did not wear their belt, most of the subjects were unwilling to change their answer, even after the interviewer clarified the intent of the question. The available response options did not allow subjects to express their behaviors adequately. A different set of response alternatives may have impacted the response process.

And finally, survey researchers need to decide when reports based on generic memories are preferable to those based on specific or episodic memories. Researchers need to consider the potential for response error depending on the wording of the question and after this decision is made, questions should be designed accordingly. In this project, the question *"On the average, how many miles would you estimate you drive during a typical week?"* demanded a considerable amount of effort for subjects to answer; most did not

answer accurately and readily acknowledged a low level of confidence in their response. The issue here is not whether the respondent is able to retrieve that sort of specific information when cued properly. Rather, the issue is whether or not the researcher needs this kind of specific information for analysis purposes. If so, is the researcher willing to either accept the response error or develop a longer series of questions which decomposes the different sorts of driving in order to achieve an accurate answer? Or will the survey objectives be met with a more general measure such as a question which would ask *"On the average, would you say you drive more than 300 miles a week, or do you drive 300 miles a week or less?"* These are but a few of the research questions we will try to address in future projects.

References

Bercini, D. (1991). "Techniques for Evaluating the Questionnaire Draft." Seminar on Quality of Federal Data, Statistical Policy Working Paper 20, Office of Management and Budget, Washington, D.C.

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

Jobe, J., Tourangeau, R., Smith, A (1993). "Contributions of survey research to the understanding of memory." Applied Cognitive Psychology, 7, 567-584.

Lessler, H., Tourangeau, R., & Salter, W. (1989). "Questionnaire Design Research in the Cognitive Research Laboratory." Vital and Health Statistics, 6(1), Hyattsville, MD: US Department of Health and Human Services.

National Highway Traffic Safety Administration, National Center for Statistics and Analysis (in press).

Tourangeau, R. (1984). "Cognitive Sciences and Survey Methods." In Jabine, T., Loftus, E., Straf, M., Tanur, J., and Tourangeau, R., eds., Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines. Washington, D.C.: National Academy of Science.

Willis, G., Royston, P., & Bercini, D. (1991). "The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires." Applied Cognitive Psychology, 5, 251-267.