Discussion, Advances in Pretesting

Floyd J. Fowler, Jr., University of Massachusetts at Boston
100 Morrissey Boulevard, Boston, MA 02125

There are three standards that good survey questions should meet: A. Questions should be consistently understood by respondents, B. Respondents should be able to answer the questions, C. (if an interviewer is used) the questions should be a reasonable protocol for a standardized question-and-answer process.

The proceedings of a conference, published in 1984 (Jabine, et. al. 1984) at which cognitive researchers and survey researchers met to discuss their mutual interests focussed on the fact that steps were not routinely taken to find out in advance if questions asked by survey researchers were meeting these standards. It is almost ten years since the Jabine report was published. To those who say things do not change, we have evidence here today that they are wrong. Things do change.

Less than five years ago, the Bureau of the Census, which is internationally known for its methodological rigor with respect to sampling, had no standards or consistent expectations for pre-survey evaluation of questions. The Bureau frequently did surveys using survey instruments that had never been pretested, much less subjected to cognitive evaluation.

At that time, there were two specific barriers that made it difficult to carry out extensive pre-survey evaluations. First, the Office of Management and Budget had rules that virtually precluded adequate pretesting. Second, the field staff at the Census was not set up and seemed not prepared to engage in the flexible kinds of question evaluation activities, including tape recording and cognitive testing, that are called for in the papers discussed this morning.

Five years ago a session such as this would have been devoted to whether or not question evaluation was important and necessary. This morning's papers all start from the premise that there will be question evaluation. The issue is what kinds of testing are best.

The paper by Davis and DeMaio is focussed on the issue of whether think-aloud procedures, that are common in cognitive testing, produce results that generalize well to actual field surveys. The approach used in the paper to explore this question is innovative. The fact, as noted in the paper, that the "control" question with which results were compared was itself more cognitively oriented than the typical survey question will, no doubt, lead to further study of the issue. Meanwhile, it is somewhat reassuring to know that think-aloud procedures per se do not distort results.

The paper by Von Thorn and Moore reports on using anthropologists to explore vocabulary and underlying issues that may affect the design of survey questions. The lack of comparability among the various anthropologic explorations reported, and the unsystematic nature of the results of their studies, do make it harder to use them. Nonetheless, that problem is common to other strategies, such as focus groups and cognitive interviews, that are used for presurvey question evaluation. The fact that anthropologists are trained to look at semantic and cultural differences among people makes them prime candidates to make a contribution to the survey process. The fact that this first effort suggests a need to further refine and structure their efforts does not imply that anthropologists do not have a role to play in the survey research enterprise.

Forsyth, Kennedy et al. provide a good, concrete example of how valuable cognitive testing can be. However, it is interesting to note that the value of their research did not lie in the fact that they uncovered subtle problems that only a skilled, cognitive psychologist could

identify. Rather, what they found was fairly straightforward and likely to be well known to any survey methodologist:

a. It is important to define a reference period so that the task is clear and consistently understood by all people

b. Do not ask two questions at once.

c. Response categories should be exhaustive and mutually exclusive.

If the principles that were violated are not remarkable or obscure, the fact still remains that had they not done the kind of testing that they did, they would not have realized that these problems existed when the protocol was developed. The key point is that the questions were subjected to critical scrutiny using techniques that can identify question problems. When such techniques are used, problems are identified and survey procedures can be improved.

Finally, DeMaio provides an excellent description of the various question evaluation options. One of the most exciting parts of the paper is that they are presented as real, cost-effective options that should be part of the Census survey process. It is also cause for celebration that some old hurdles, previously posed by the Office of Management & Budget are no longer in place, at least for the Bureau of the Census. Moreover, the field staff at the Bureau of the Census clearly has proven itself capable of participating in flexible, useful question evaluation processes.

Having said that this session is a time for feeling good, because of the growing maturity and acceptance of the value of good question evaluation, there are at least three cautionary points on which I want to close.

First, the battle certainly is not won. All researchers do not believe that cognitive testing, or some variation thereon, is an essential part of the survey instrument design process. In fact, a paper given at this conference by Johnny Blair reported that only a third of all academic survey organizations have ever used cognitive testing techniques to evaluate their survey questions.

Second, while it is encouraging that the Bureau of the Census can produce such a good summary of question evaluation options, it is still not clear that there is an organizational commitment to make such evaluation procedures a routine part of the development of all surveys. I have no quarrel with DeMaio's contention that question evaluation protocols need to be flexible and can appropriately vary from survey to survey. Nonetheless, we also need agreement that some minimal protocol for question evaluation should be part of any survey that bears the stamp of approval of the Bureau of the Census. To date, there are no such standards at the Bureau of the Census, or anywhere else of which I am aware. Until those standards are in place, so that all survey sponsors assume that their budgets and their time schedules have to allow for such question evaluation activities, they will frequently be omitted, and the quality of the resulting data will be worse for it.

Finally, we know that when survey questions are tested, and problems are found with them, sometimes they are asked anyway. There are at least two reasons for this. First, our standards for when a possible problem indicated from cognitive evaluation or pretesting is indeed a problem are not very clear. We need to work on how to make the results of question evaluation more quantitative and more consistent. Second, some survey questions have a long pedigree; one of the reasons people want to ask questions is to compare them with previous results. Under those conditions, sometimes researchers choose to ask questions even though they have been demonstrated to have major problems from the point of view of being good, standardized survey measures. We probably need to preserve the right of researchers to ask bad questions. However, if problems have been clearly identified during presurvey testing, those

problems should also be described fully in the methodological appendix reporting the results. To date, we have standards for reporting response rates and sampling errors. However, we do not have either standards or expectations for reporting results of presurvey evaluations of questions. When researchers are expected to report the results of their cognitive questions, and a standard part of a methodological appendix is how well questions stand up to cognitive testing and behavior coding during pretesting, we can expect researchers to take the results much more seriously, as well as to have better informed readers.

In conclusion, there is still work to be done. Nonetheless, the fact of this session and these four excellent papers focussed on how best to do presurvey evaluation of questions is encouraging testimony to an important area of progress in improving survey research methods.

Reference

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

GE15 (ASA)