

Can Cognitive Information Be Collected Through the Mail? Comparing Cognitive Data Collected in Written versus Verbal Format

by Wendy Davis, Theresa J. DeMaio and Andrew Zukerberg

Key Words: Establishment surveys, pretesting methods

INTRODUCTION

Cognitive pretesting of questionnaires is quickly gaining recognition as a useful way for identifying possible causes of survey response error (Tourangeau, 1984; Forsyth & Lessler, 1991; Willis, Royston & Bercini, 1991). Specifically, these techniques are used to help identify problems with comprehension of the survey items as a whole as well as specific terms, understanding and use of the response scales, and so on. Though there are many cognitive methods which can be used for pretesting, a very frequently used method is some version of a cognitive interview. Concurrent think aloud interviews, retrospective interviews with probing, interviews using paraphrasing techniques, vignette classification, as well as others, are examples of techniques used in conducting cognitive interviews. The common element is that they all involve respondent participation and are interviewer-administered.

Each of these techniques is useful in identifying problems in one or more components of the response process. However, because they are interviewer dependent, they are difficult to adapt to the response process for surveys of establishments. In fact, despite what these methods reveal about the response process, Christianson and Tortora (1995) report that in a survey they conducted concerning the number, type and content of the establishment surveys conducted by various organizations, there were no plans to incorporate such cognitive techniques in any future redesign efforts.

Given that cognitive techniques seem to be accepted as useful methods for pretesting and redesigning questionnaires, why are these techniques not used more for establishment surveys? First of all, it is our opinion that establishment surveys are generally not pretested as frequently as household surveys. Secondly because cognitive techniques are interviewer-administered, they are less suitable for establishment surveys for several reasons (Edwards & Cantor, 1991). First, in an establishment survey it is typically necessary to access the knowledge or expertise of more than one person within the company. However, it may not be possible to gather all of the individuals together at once in order to conduct a cognitive interview, especially in large companies with complex structures.

Secondly, in many instances the information requested on the survey form is not information that respondents have in memory, so they have to use

records to respond (Cox et al., 1989). Identifying and locating the appropriate records can be a time-consuming process that cannot always be completed successfully in the time period when an interviewer can be present. Third, for many establishment surveys, estimated reporting burden is quite high -- in the range of 10-15 hours or more. Clearly, an interviewer cannot be present for the entire period of time it takes to complete the form and a respondent is not likely to complete a form of that length in a single setting. Lastly, though not unique to establishment surveys, it is costly to send an interviewer to numerous locations across the country.

There is not an abundance of pretesting for establishment surveys reported in the literature, especially those using any type of cognitive technique. The available literature uses the techniques of retrospective structured telephone interviews (Phipps, 1990); retrospective personal interviews (Phipps, 1990; Palmisano, 1988); in-depth and think aloud interviews (Gower and Nargundkar, 1991; Palmisano, 1988); and focus groups (Gower and Nargundkar, 1991). These techniques have been reported to yield different kinds of information about the response process. For example, Phipps (1990) reports that structured phone interviews mainly indicate errors of omission whereas personal interviews indicate comprehension problems. However, it was noted by some of these authors that a common problem with the personal interview was a lack of availability of the necessary data for providing answers to the survey items within the time period of the interview.

This paper reports the results of a questionnaire redesign effort for an establishment survey using traditional cognitive interviews, as well as a second method which did not require the presence of an interviewer. The second method was a mail-out/mail-back questionnaire with cognitive probe questions embedded directly onto the report form. The focus of this paper is to compare the content of the cognitive information collected using the two techniques. The paper is organized as follows: first we describe our research study design, focusing on the development of the cognitive mail form. In the next section we compare the results of the cognitive interviews with the results of the mail-out/mail-back study. Finally, we present some suggestions for ways to improve the mail methodology used in this study and discuss the advantages and limitations of the mail technique.

OVERALL STUDY DESIGN

This research was conducted on the Survey of Industrial Research and Development, conducted by the U.S. Bureau of the Census for the National Science Foundation. This survey is designed to measure the nature and support of corporate research and development. There are two versions of the form: short forms administered annually and long forms administered biennially. Companies known to have research and development expenditures of one million dollars or greater are selected with certainty and mailed the long form. Short form recipients have less than one million dollars in R&D expenditures and their probability of selection is less than one. The level of R&D expenditure is determined either from past responses to this survey or from secondary sources. Our research followed these same rules for determining who would receive our research versions of the long and short forms for the mail portion of the study.

Development of the mail-out/mail-back form

Our overall objective when designing the mail-out/mail-back form was to collect the same information using a paper-and-pencil format as we collect in a personal visit interview. Thus, the questions we developed for the mail-out/mail-back forms shared the same focus as the cognitive interviews. We refer to the questions developed for the mail forms as Cognitively-Oriented Debriefing (COD) questions, all of which were open ended. The COD questions were embedded directly onto the report form immediately after the survey item to which they pertained.

The COD questions, though identical in focus to the cognitive probes, are worded in a slightly more directive manner. The more directive nature of the COD questions is based on the "thought-listing" literature taken from the field of cognitive assessment. Cognitive psychologists (Blackwell et al, 1985) find that when requesting people to write rather than verbalize their thoughts, people are apt to record things that are abstract in content instead of recording the actual mechanics of their response formulations. They rarely volunteer the strategies or procedures they use when forming their response. For our purposes the strategies used to develop a particular response are important pieces of data for diagnosing errors in the response process, so the COD questions we developed were intended to lead respondents to write about their mechanical-strategic cognitive processes. For example, in reference to an FTE figure for company scientists and engineers, rather than simply asking respondents "how did you come up with this value for scientists and engineers?" as we did in the personal visits, we made a slightly more direct request by saying "Please give a detailed explanation of the method you used to come up

with the number of scientists and engineers you reported in item 5."

Another difference between the two methods is that in a think aloud interview it is often unnecessary to ask a specific probe because the respondent volunteered the sought information. On the other hand, in the written format, without the particular stimulus (COD) question the respondent is not likely to provide the cognitive information at all.

Respondent Selection and Procedures

Our total "sample" size was 100 for the mail-out/mail-back portion of the study, though we were expecting only slightly more than half to agree to participate. The 100 companies were geographically distributed across the United States. Twenty-five companies were selected for the research version of the long form with inserted COD questions. The remaining 75 companies were selected for the research version of the short form with inserted COD questions.

Forty companies were selected to participate in the personal visit portion of our cognitive research using the short form. All forty companies were known to have research and development and they were selected from the Northeast corridor to contain travel costs. Of these 40 cases we expected to get 15 personal interviews.

In each condition, companies were contacted by phone initially to solicit their participation as well as to identify the most appropriate respondent.

In the mail-out/mail-back condition, if respondents did not agree over the phone to participate, they were not mailed a form. In addition, for OMB clearance reasons we were only able to mail 48 short forms and 23 long forms. Once we reached this number, we discontinued contacting any more companies for the mail-out/mail-back condition. Each company that did agree to participate was mailed a research questionnaire with a cover letter that restated the purpose of our research, and explained how to complete the report form with the inserted COD questions. Mail-out/mail-back companies were each given approximately a month from the day we expected them to receive the form to complete it. Follow-up calls were made about 2 weeks after mail out if the form had not been returned and again 1 and 2 weeks later if the form still had not been returned.

ANALYSIS PROCEDURES

Content analysis was done on the written responses to the COD questions and on the verbatim transcripts of the cognitive interviews. Personal visit respondents were only probed about the core items which had COD items included on the research versions of both forms. The coding scheme consisted

of 3 categories of codes, each of which reflected different aspects of response formation: 1) codes that captured what dimensions respondents thought about as they formed their responses (completeness codes); 2) codes that captured the sources of information, such as records or personal knowledge, respondents used to form their answers (source codes); and 3) codes that captured the content of their responses, such as their interpretation of terms and reference periods (content codes).

Due to time and staff limitations, only 2 of the core set of survey items were coded. The survey items selected for coding were the two most central to the survey's data collection effort -- research and development expenditures by specific type of research, and the full time equivalent (FTE) number of scientists and engineers doing company research and development work. Each mail-out/mail-back form, and the transcript of each cognitive interview was coded independently by two coders and all differences were reconciled. The overall percent agreement before reconciliation was 83%.

RESULTS

Level of Cooperation

Cooperation is difficult to measure since we did not use a probability sample, and since cooperation was first solicited over the phone for both modes of data collection. Given this, cooperation is defined as encompassing those people who actually completed at least some part of the form once they had already agreed over the phone to participate in our research. In the personal visit portion of the study, we were expecting to get 15 interviews from the 40 cases in the sample. We were only able to conduct 11 interviews.

In the mail-out/mail-back portion of the study, 48 forms were mailed to short form companies of which 40 were returned (83%), and 23 were mailed to long form companies of which 20 were returned (87%). Of the 20 companies that returned the long form, 3 returned the form without answering any of the COD questions. All short form companies who returned the form completed at least some subset of the COD questions.

Methodological Results

Completeness Codes

The completeness codes are the least specific level of evaluation. Basically, these codes tell us whether there is enough information to identify what respondents thought about when responding, regardless of correctness. There are 3 completeness code variables for each of the two survey items. Each of the variables reflects a critical dimension of the survey item as identified by ourselves and NSF staff. For the scientists and engineers question, the critical dimensions include whether the respondent considered the

education/experience criteria, mentioned the FTE calculation, and noted the reference period used when providing their response to the survey item. For the research and development expenditures questions, the critical dimensions of the completeness code variables cover whether the respondent addressed the issue of new versus existing knowledge, and specificity to particular products or processes. These codes apply to the three types of research: basic, applied and development.

The general trend when examining the completeness codes is that the personal visit respondents provide the most complete information of the three types of respondents. For example, approximately 91% of the personal visit respondents provided complete information about their responses to the FTE number of scientists and engineers in their company. On the other hand, only 59% and 56% of the short and long form mail respondents respectively provided the equivalent information.

Each of the completeness variables is measured at three levels: 1) the respondent explained the critical dimension clearly, 2) it was unclear how the respondent thought about the critical dimension, or 3) the respondent did not address the dimension at all. The other systematic trend in the completeness codes is reflected in these different levels of the completeness code variables. For both survey items, the long form respondents were somewhat less likely to address the critical dimensions of the questions. For example, in defining the types of research performed by the company (i.e., basic, applied or development) the respondents have to distinguish whether the objective of the research is to pursue new knowledge, or if the research is based on existing knowledge and has commercial objectives. Research in pursuit of new knowledge is "basic" research, but research that primarily uses existing knowledge for meeting commercial objectives is either applied or development work. Thus, a critical dimension of the R&D expenditures survey item is whether they told us what the knowledge base was for the research they report. Less than half of the respondents in each of the mail conditions provided this information, whereas 67% of the personal visit respondents provided the equivalent information. This same trend was observed in the scientists and engineers question.

This trend is not replicated, however, for another critical dimension of the R&D expenditure question -- whether the reported research is product or process specific. If it is product or process specific, it can't be basic research. Thus, in order to determine if the reported research has been classified correctly we need to know this information. The personal visit

respondents provided this information less frequently (about 72% of the time) than did the mail respondents (about 81% and 85% for the short and long form respondents, respectively). This is most likely due to the kinds of responses given in the written format. Mail respondents were more apt to talk about their research in terms of the product. For example, a typical mail response explaining their company's research was "testing water-based inks on film substrate which has not been tested or applied in the printing industry." Primarily, we think this reflects that respondents didn't always know the products that mapped to the research expense line recorded in their financial records, and in the time of a personal interview, they were not always able to find someone who could give them this information. Mail respondents, on the other hand, had more time available to wait for a response back from a more knowledgeable colleague.

In summary, the results of the completeness codes suggest that a mode of administration effect may exist, and that mail respondents are less likely to provide complete information explaining the dimensions they considered when developing their response. However, if specific information is requested that is not a necessary part of the record keeping system (e.g., product information), personal visit respondents may not be able to provide this information during the course of an interview since they have to rely on outside help. In this instance mail respondents do better.

Source Codes

The source codes for the scientists and engineers question and the research and development expenditures question were identical. They captured whether the respondent identified the information as coming from their personal knowledge, from a colleague, or from records. The main trend in the codes capturing the source of information is that across both the scientists and engineers question and the R&D expenditure question, the personal visit respondents provided a great deal more information than either of the two mail conditions. Only about 10% of mail respondents provided source information whereas all personal visit respondents did. Given the differences in the setting between a personal interaction involving two or more people and a self-administered instrument, these differences between the reported procedures is not surprising. When a cognitive interviewer is in the room with the respondent, norms of politeness lead him/her to indicate why they are leaving the room (i.e., to get records stored elsewhere), or why they are going to make a phone call while the interviewer is waiting (i.e., to ask a colleague a question.)

This difference may even have been compounded by the way in which we obtained this information on

the mail forms. As a measure of the number of people involved in completing the item we asked respondents to record the initials of all people who took part in answering a survey item. This request did not require the respondent to give us details about mode of contact with colleagues or colleagues' titles, both of which were volunteered in a personal visit setting. But by providing the initials of people who assisted in completing the form, respondents in the mail condition believed they had provided all the necessary information and had met our request.

Content Codes

The content codes captured information about how respondents interpreted the questions and terms within the questions. For example, we wanted to be able to tell if people were able to understand the phrase "equivalent" to a four year degree and include these people in their response. We only used content codes for the scientists and engineers question because the content of the research and development expenditure questions was quite complicated and time did not allow for us to develop a coding scheme for that level of complexity.

There were 3 different substantive codes for the definition of a scientist and engineer: someone with a 4 year degree in a relevant field, or the equivalent; only people with 4 year degrees; or anyone on the R&D staff. The response with the highest percentage reported in the mail study was "4 year degree" only, whereas it was "4 year degree and the equivalent," which was the correct response, in the personal visit study. There were 4 different substantive codes to describe the method used for calculating FTE's: the correct FTE procedure as defined by NSF; FTE's based on an arbitrary proportion; a head count only; or any other procedure. The highest percentage of responses for the mail respondents was "head count only," whereas the highest percentage of responses for personal visit respondents was "FTE based on an arbitrary proportion." (Actually, neither one of these responses is definitively correct but the response given by personal visit respondents has a higher probability of being correct.) Thus, there does seem to be a systematic difference between the written mail responses and those given verbally in the interview.

This difference may reflect a true difference in the procedures that mail respondents as compared to personal visit respondents use for answering the survey item, or it may reflect a difference in the amount of effort respondents are willing to expend to report in the COD questions the procedures they used to answer the survey item. In the case of the mail-out/mail-back respondents, regardless of what number they actually reported as their survey response, their

response to the COD items may reflect that it was psychologically easier to write that they did "a count of the people in the R&D department" (e.g., a head count) rather than to say, for example, that they determined some percentage of time people spent on R&D in the past year, and multiplied that percentage by the number of people who worked on R&D to get a total. This would be the reporting procedure for an FTE based on an (arbitrary) proportion.

To try to decipher what the difference truly reflected we looked at the number of fractional values reported in response to the survey item. Assuming fractional values are more likely in FTE values than in head counts, one would expect that the number of fractional values reported for this survey item would be greater for personal visit cases than for mail respondents if the mail respondents truly are answering with a head count procedure only. On the other hand, one would expect the number of fractional numbers reported for the survey item to be roughly equivalent for the mail and personal visit cases if the difference in reported procedures were actually reflecting the mail respondents unwillingness to expend the effort to explain the more complicated procedure in a written response.

As it turns out, about 18% of both the mail and personal visit companies reported fractional values to this survey item. This suggests that a mode effect may exist. Even though short form respondents in the mail condition reported an equal number of fractional values as their personal visit counterparts, their written responses explaining the numbers they reported did not reflect this. This supports the notion that the respondents may simply be recording an easier, less detailed response to the COD questions just to get an answer down without expending a lot of effort. This undermines the usefulness of this method for detecting the source of the problem within a given survey item.

DISCUSSION

In general, across all categories of codes, and across both questions, the personal visit respondents consistently provided more complete and detailed information than did the short or long form respondents in the mail condition. This suggests a mode of administration effect, though this inference must be tempered given the differences between "samples" in the three conditions. A difference in responses is especially evident in the analysis of the source codes. In general, short and long form respondents in the mail condition both performed quite poorly as compared to the personal visit respondents in terms of providing clear and detailed data about the information sources they used to develop their survey response. In fact the percent of "no mentions" is the overall modal category for the mail respondents, suggesting that they were not aware that

this information should be reported. Without a specific COD item requesting this information about their record source, or specific instructions informing respondents to report this information, it is not surprising that they did not. Another difference is in the modal categories of the content codes. Respondents to the mail form described less complicated processes, even though based on their survey answers it seemed that they did something more comprehensive than what they described.

There are several possible reasons why these findings are not surprising, though admittedly disappointing. First, in the personal visit mode, nonverbal gestures such as an expectant stare at the end of a response or carefully watching the respondents' movements can provide respondents with feedback for evaluating their responses to the cognitive probes. This same feedback is not possible with a self-administered instrument.

A second but related difference between the two modes is that the interviewer is given the opportunity to restate a probe that seems to have been misinterpreted by respondents. On the other hand, in the mail mode, the researcher has one chance to make the intent of each probe or COD question clear. Thus, as suggested by the high proportion of "unclear" and irrelevant responses as compared to the substantive responses, respondents didn't always accurately interpret what they should be writing in response to the COD items.

A third major difference between modes is that there are norms governing social interactions which don't exist in a self-administered instrument. In a personal visit mode, the respondent is likely to volunteer information about colleagues they contact for assistance, or records they have to search for, because to do either of these things, the respondent has to interrupt the interaction between him/herself and the interviewer. In a self-administered form, there is no such interruption to explain.

A final difference between the two modes that can account for the discrepancy in the responses to the COD and personal visit probes is the greater effort involved in writing responses to the COD items as opposed to verbalizing this same information.

All of these differences suggest that the mail-out/mail-back cognitive technique could be designed differently to make it more comparable to the personal visit technique. To get more information about the dimensions considered in giving a response, techniques can be systematically introduced into the mail design that try to match the more comprehensive nature of the personal visit probes. For example, as noted above, norms of social interaction which are

only applicable in the interview mode, might be brought into a mail form by adding language to the form that appeals to these same norms. Respondents could be instructed to answer as if a (naive) researcher was present who needs every detail explained. Work by Cannell et al. (1981), shows that providing respondents with instructions for completing the task improved the quality of the data they provided.

Another technique related to this same work by Cannell et al., which would make the two modes more comparable is to add a "practice question" at the beginning of mail form. In a think aloud personal interview, the respondent is given a "practice" question either by the interviewer providing a model response to a cognitive probe, or having the respondent give the response and the interviewer providing feedback. This same technique could be incorporated into the mail mode by providing a hypothetical answer to a practice COD item which could then serve as the model for what level of detail respondents should provide in their answers to the COD items.

There are several advantages to using a mail technique for collecting cognitive information. First, the cognitive pre-test is done in the same mode of administration as the actual survey itself. Thus, respondents can use the same strategies they use to respond to the actual survey, rather than adapting their strategies to an interviewer administered task.

Another advantage is that the cost per unit response is reduced and travel costs involved with visiting different establishments are minimized or eliminated. As a result larger samples can be used. In addition, the sample can be selected according to the same design as the actual survey (i.e. multi-stage probability sampling) to make the results more generalizable.

A final positive point is that even though the personal visit information may have been more complete, the mail method still provided useful data. With larger sample sizes and improved COD questions this method could potentially yield a lot of data on which to base a (re)design effort.

CONCLUSIONS

The ultimate goal of this work was to investigate whether a self-administered mail method was a viable alternative to cognitive interviews for the purpose of questionnaire redesign and evaluation. Given the results presented here, our decision is yet to be made. First, there were several problems with the design of the mail instrument that we believe impeded the effectiveness of the method. Second, we have not yet done the most convincing test. A better test would involve strict experimental controls with random assignment of cases to the personal visit and mail conditions. Then 2 questionnaires could be developed independently, one

based only on the mail cognitive data, and the other based only on the personal visit data. The two redesigned questionnaires would then be compared for consistency. Thus, our conclusions are simply the inevitable call for more research.

References

- Blackwell, R. T., Galassi, J. P., Galassi, M. D., & Watson, T. E. (1985) "Are cognitive assessment methods equal? A comparison of think aloud and thought listing." Cognitive Therapy & Research, 9
- Cannell, C., Miller, P., and Oksenberg, L. (1981) "Research on Interviewing Techniques" in S. Leinhardt (ed) Sociological Methodology, San Francisco: Josey-Bass Publishers
- Christianson, A. and Tortora, R. (1995) "Issues in surveying businesses: An international survey" in B. Cox, D. Binder, B. N. Chinnappa, A. Christianson, M. Colledge, and P. Kott (eds.) Business Survey Methods, John Wiley & Sons, Inc., chapter 14.
- Cox, B. C., Elliehausen, G. E., and Wolken, J. D. (1989) "Surveying small businesses about their finances" Section on Survey Research Methods, American Statistical Association, pp 553-557
- Edwards, W. S. and Cantor, D. (1991) "Towards a response model in establishment surveys" in Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.) Measurement Errors in Surveys, John Wiley & Sons, Inc., chapter 12
- Forsyth, B. H., and Lessler, J. T. (1991) "Cognitive laboratory methods: A taxonomy" in Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.) Measurement Errors in Surveys, John Wiley & Sons, Inc., chapter 20
- Gower, A. R., and Nargundkar, M. S. (1991) "Cognitive Aspects of Questionnaire Design: Business Surveys versus Household Surveys" Proceedings of the 1991 Annual Research Conference, Washington DC: Bureau of the Census
- Palamisano, M. (1988) "The application of cognitive survey methodology to an establishment survey field test" Proceedings of the Section on Survey Research Methods, American Statistical Assoc., pp. 179-184
- Phipps, Polly (1990) Proceedings of the Annual Meeting of the American Statistical Society, Section on Survey Methods Research
- Tourangeau, R. (1984) "Cognitive Science and Survey Methods" in Jabine, Straf, Tanur, and Tourangeau (eds.) Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, National Academy Press, Washington DC
- Willis, G., Royston, P., and Bercini, D. (1991) "The use of verbal report methods in the development and testing of survey questionnaires" Applied Cognitive Psychology, 5, 251-267