

## Overcoming Structure and Order Effects in Cognitive Interviewing

Justin T. Bailey<sup>1</sup>, Kenneth W. Steve, Anh Thu Burks, Kimberly D. Brown, and Paul J. Lavrakas  
Nielsen Media Research

### Abstract

Much research has been done to determine the most effective probing and presentation methods during cognitive interviewing. However, there are times when it is unclear which methods will be most effective. Nielsen Media Research (NMR) recently conducted a series of cognitive interviews which revealed several insights into how to avoid common pitfalls. First, it was determined that one form of cognitive interviewing prevailed over another in producing more consistent, meaningful responses from participants. Specifically, an initial set of questions closely followed a telephone survey script being tested by NMR. When cognitive interview participants failed to give adequate “think aloud” responses, the question order was revised in attempt to increase verbalization and alleviate confusion. The second design, which utilized more “bottom up” processing techniques, resulted in meaningful responses that improved the quality of the telephone script. Additionally, during the second part of the cognitive interview, participants were asked to provide feedback on visual stimuli that were being tested for potential use in NMR’s TV Ratings Diary.

Although cognitive interviewing has become a widely-used tool for assessing questionnaire quality, using this method of interviewing for evaluating non-question items such as visual stimuli has not been examined. Participants were given two sets of designs and asked to verbalize their thoughts as they looked at each. Initial results revealed that presentation order determined which design the participant favored—they tended to choose the last set presented. However, when a third design was presented to participants, the order effects were minimized. Moreover, participants gave more specific qualitative feedback in regards to each design when presented three designs compared to only two. The implications of these findings are examined in the context of cognitive theory. These findings are further discussed as a way to improve probing techniques and question order in cognitive interviewing.

**Keywords:** Cognitive Interviewing, Cognitive Processes, Order Effects.

### 1. Introduction

#### 1.1 Cognitive Interviewing

Survey design has long been an important issue for survey methodologists. The impact of designing successful surveys is critical for producing valid, clearly-definable results. One method of obtaining successful survey questions is that of cognitive interviewing.

During cognitive interviewing, researchers pre-screen survey questions by asking participants to answer the potential questions and explain why they answered it as they did. This enables the researchers to identify latent problems with the questions that are not easily revealed. To understand what the participants are thinking, *probing* is used to gain further insights. Through more than two decades of research, many different types of probes have been identified and employed by cognitive interviewers (Willis, 2001).

Forsyth and Lessler (1991) were the first to create a taxonomy for these probes. Building on previous cognitive processes work (Ericsson & Simon, 1980), they suggested that there were two factors for effective cognitive interviewing: (1) task timing and (2) attention control. The first of these, task timing, “refers to the timing of the reporting or data-gathering task in relation to the question-answering activities we wish to study” (p. 406). Thus, the exact timing of when the probing attempts occur varies. The second, attention control, is defined as certain filters that researchers use to aid the thought processes of the respondent during the cognitive interview.

One point at which the probes are administered is concurrently with the presentation of the survey questions. The most common form of concurrent probing is the think-aloud process, in which participants are asked to express what they are thinking verbally as they answer a particular survey item. The goal of concurrent probing is to ascertain question limitations as soon as possible from when they are presented to a respondent.

Forsyth and Lessler (1991) identify a second type of task timing called immediate probing. These are probes that occur directly after the presentation of the survey items being pre-screened. Examples of this technique include paraphrasing and follow-up probes, both of which involve asking the respondent to repeat

the survey question back to the interviewer. These types of probes identify problems with the design of the survey by eliciting the respondent's understanding of a question after it has been asked and he or she has responded.

A third type of task timing is delayed probing. These are probes that occur sometime after the completion of the survey questions. They are most commonly called retrospective think-alouds, and are similar to concurrent think-alouds except that the interviewer does not ask the respondent to verbalize thoughts until after the survey has been completed. This may be effective for particular surveys in which the researchers want to convey the entire meaning of the questionnaire before probing.

There has been much debate about which of these types of probes is the most effective for assessing response errors. For example, Daugherty et al. (2001) tested three different approaches, depending on the type of the probe and the timing of the probe.

After doing cognitive interviews on the same questionnaire using three different approaches, Daugherty et al. concluded that (1) concurrent probing seemed to identify more problems with the questionnaire than did retrospective probing, (2) the concurrent, general approach appeared to be more useful for identifying issues of comprehension and judgment, and (3) concurrent, tailored probing uncovered more misunderstandings in what the person recalled about the question being asked.

All three of these findings appear logical given the purpose of cognitive interviewing. If the goal is to reveal specific problems in the questionnaire, probing directed at these issues makes the most sense. If the goal is to discover unanticipated issues with questions, then asking general, "tell me what you are thinking" probes would logically be the proper approach. Retrospective approaches tend to leave time for the respondent to lose the ability to introspect (Nisbett & Wilson, 1977). In such instances, people are unable to give an accurate verbal explanation of their cognitive process and instead rely on prior beliefs of how they think they should feel about the question. This leads to inaccurate or incomplete verbal reports. Further, Ericsson and Simon (1984, 1993) suggest that the ability to provide verbal reports retrospectively lasts only as long as it remains in our short term memory, usually only for a few seconds.

Still, others have argued that there is little difference between retrospective and concurrent probing (Redline et al., 1998). This appears to be especially true if the probing is done directly after questionnaire administration (Willis, 2001).

One unique application of the retrospective technique was utilized by Schechter and Beatty (1994). They had participants go through a simulated phone

call test script in its entirety before probing retrospectively. After the phone call was completed, the interviewer rejoined the participant to ask her/him specifics about each question, using paraphrasing and other tailored probes. They concluded that this technique was effective in this situation because it enabled interviewers to observe the flow and identify hesitations throughout the interview. Further, they hypothesized that the participants could clearly separate the two cognitive tasks of responding to survey questions and responding to probes.

From a theoretical standpoint, these different forms of probing could rationally be examined in terms of cognitive processing styles. As Tourangeau (1984) pointed out in his seminal cognitive interviewing framework, people use both top-down and bottom-up styles to process information. In the top-down view, we use what we already know to make sense of a question. Therefore, we have to have an a priori concept of the terms in the question in order to accurately understand it. Cognitive interviewing can unearth one's preconceived ideas about the subject of the question so that the interviewer can understand any miscomprehension.

As for the bottom-up view, we build an image of what is being asked through a sort of piecemeal fashion. That is, we obtain information about what is being asked by listening and developing a broad concept in our mind. Cognitive interviewing can reveal what pieces are important for one to draw a conclusion about what is being asked. If one of the pieces is misunderstood, it could potentially be revealed through probing about the specifics of the question. However, what has not yet been thoroughly researched is the interplay between these two complementary processing styles in respect to cognitive interviewing. Are certain approaches more likely to lead a respondent to use one processing style over the other?

For example, it could be that top-down processing is used during retrospective probing and bottom-up processing during concurrent probing. Or that general probes lead a person to build their own concept from the bottom-up while tailored probes require that a person start with a pre-existing concept and work down. The implications of this could affect the design of cognitive interviews, based on what is desired from the outcome.

In the current research, we will look at this in greater detail. One purpose of this paper is to bring to light this theoretical possibility; that is, to revisit the notion that different cognitive interviewing approaches can lead a respondent to use various processing styles. In investigating this possibility, the goal of this paper is not to test this hypothesis through experimentation, but rather to explore and generate hypotheses for

future researching into cognitive interviewing methods. Nor is the goal to state that one approach is better than another, but rather to suggest that the types of probes an interviewer uses could potentially influence how the respondent thinks about a question.

### 1.2 Order Effects

The typical method for cognitive interviews is to present a set of questions that have been determined in advance. Either during or after the presentation of the questions, an interviewer probes the respondent as to how he or she arrived at the answer (Willis, 2001).

However, what about instances when the goal of the interview is to assess written instructions and graphics as opposed to actual survey questions? What if a person is asked to think aloud about two different instructional designs presented to her/him? As will be discussed in greater detail later in this paper, presenting these types of instructions to respondents can be both informative and difficult using cognitive interviews. If the interviewer can overcome the tendency of the respondent to like what has most recently been presented to her/him, it is possible to learn about potential flaws in one instructional design versus others. Through this research, we hope to speculate about using cognitive interviews for this type of design framework.

## 2. Background

As a way to measure the television viewing patterns of households nationwide, NMR conducts "sweeps" eight times a year. Randomly selected households are recruited to participate via telephone and sent a Nielsen TV Diary to record their viewing for a one-week period. Recently, NMR has been forced to adapt to the rapidly-changing technologies that shape the way television is viewed. To this end, NMR began using cognitive interviews in 2002 in order to better understand how well people comprehend the questions asked of them and the instructions posed to them. Specifically, members of NMR's methodological research department explored how survey participants respond to questions typically given to them both on the telephone and in written form.

The first round of cognitive interviewing on the topic that is reported in this paper was conducted in September 2004. Based on these interviews, NMR modified its verbiage for a Digital Video Recorder question so that those households could respond accurately whether or not they had this technology. A second round of cognitive interviews was built on this first round, and tested questions about both digital cable and On Demand.

## 3. Phase 1 - Method

In October and November of 2005, NMR's methodological research department conducted a series of cognitive interviews with the goal of improving yet untested questions regarding two new technologies that are integral parts of television viewing: Digital Cable and On Demand. Additionally, we sought to learn more about our questions that asked about a Digital Video Recorder (DVR).

The interviews were conducted with 33 former NMR diary-keepers Tampa, Baltimore, and Austin. These three cities were selected based on differences in their viewing options and viewing trends in the diaries of sampled homes. Respondents were previous diary-keepers so that we knew in advance what types of technology they had. The goal was to interview a broad array of people who may or may not be familiar with the types of technology being tested. It was hoped this would provide us with the most wide-ranging number and type of responses. Each interview lasted between 60 and 90 minutes, and each participant was given \$100 in cash as a thank you for the interview. The interview consisted of two main parts.

### 3.1 Telephone Script

For each of the diary cycles conducted by NMR, households are recruited via telephone. Each household is called and given a similar script. Based on the answers to the script, we know which diary to send to the household. Households with more advanced technology receive an 8-day diary that enables them to enter in programs that have been "time-shifted". For those individuals without this technology, a more basic 7-day diary is mailed. Thus, making sure that people understand the questions during the phone script is highly important to ensure that households receive the correct diary.

Therefore, the first part of our interview focused on new technology questions that we were pre-testing for the phone script. We initially designed the script similarly to Schechter and Beatty's (1994) cognitive interviews of telephone surveys with the following exceptions. First, we did not completely simulate a telephone call in a laboratory. Instead, we conducted the standard face-to-face interview while reading through the phone script. Secondly, because many of the questions we have in our phone script are not germane to the topic of the questions we were testing, we created an abbreviated version that focused on this part of the script. The end result was 8-10 questions that were being tested for use in the recruitment script.

These questions were read verbatim to each participant and probing did not begin until after all

questions were read. Participants were encouraged to elaborate on their responses, although this was not required. At the completion of the last question, a set of tailored probes were asked to assess participants' understanding of the questions that were read to them and to learn their general understanding of the technologies in question. Participants were asked to paraphrase the questions and describe each of the types of technology in their own words. The interviewers followed up with other probes as needed.

### 3.2 Diary Design

After each participant had given thorough responses to each of the script questions and probes, the interviewer then moved on to the second part, which was to learn about the actual visual design of the Nielsen TV Diary. After some initial warm-ups, participants were handed a diary and instructed to go through each page as they would if they had just received it in the mail, and to continue "thinking out loud" about what they were looking at and thoughts they might have.

For the purposes of the test, only one of four steps of instructions was altered. This step included new information for entering in programming that was watched with a DVR or On Demand. Each participant was randomly presented two of the eight test designs for this step. Each design differed slightly from the others on some important aspect, so that differences could be compared. The presentation order was varied so as not to introduce an order effect. On the initial part of the general think-aloud, the interviewer did not draw attention to this step or ask specifically about it. The objective was to learn the initial reactions of the participants.

Participants would typically look at each page and make comments as they read over the instructions presented to them. When the participant indicated that he or she was now ready to begin entering programs, the interviewer stopped the think-aloud procedure and informed the participant that he or she going to be read some scenarios and for them to enter the scenarios into the diary while thinking out loud. This, coupled with the participants' think-aloud responses, were the measures for success on this particular task.

At the completion of all seven scenarios, the interviewer removed the diary and presented a new version, open to the modified step in the instructions. The participant was, as usual, asked to think aloud about the design of the second diary. This was followed by more specific probing about specific parts of this step. Finally, participants were presented with both diaries and asked to elaborate on which they liked the best, which they found easiest to use, etc. The thoughts participants gave on this were important for determining the best diary to use in the future.

## 4. Phase 1 - Results

### 4.1 Telephone Script

Since our aim at NMR was to test new questions that would be used for recruitment via telephone, we hoped that the cognitive interviews would provide insight into how people understood the questions. In order to have the highest quality data, it is critical for us to be clear on what we are asking during recruitment. The belief was that getting people to paraphrase what they believed us be asking would identify concerns that they might have. Paraphrasing is a common technique used by cognitive interviewers to elicit how well a person comprehended a question (Forsyth and Lessler, 1991; Martin, 2004). Hess and Singer (1995), among others, found that paraphrasing was an effective way to pretest complex questions. Given the level of complexity involved in our questions, a logical approach was the use of paraphrasing.

However, the initial feedback we received from the interviews was evidence to the contrary. Instead of learning new information about our questions, our respondents gave us relatively little information about the question wording. Of the 11 people with which we used this cognitive approach, we received an average of 5.0 repeated items on the four questions we asked them to paraphrase. There was a possibility of paraphrasing 17 such key points mentioned in the four questions. Thus, for each question, participants were only able to repeat 1.25 of the key concepts, out of a possible 4.25. Additionally, we also coded new items suggested by participants at any time during the interview. A suggestion was coded if it modified the current definition in any way, even if the suggestion was inaccurate<sup>1</sup>. For these, our first 11 participants averaged a total of 4.63 additional comments or suggestions (1.15 per question). Thus, not only were participants unable (or unwilling) to repeat back what the interviewer had just read to them, they also failed to suggest new ideas that would improve the quality of the questions being asked.

It is important to note that the above analysis was done post hoc, after all interviews had been administered and coded. The research team was not privy to this information at the time of the interviews. However, it was agreed upon that the approach being used in the interviews was not fruitful in bearing the information desired. It was determined that the best measures at the time were to change the cognitive

<sup>1</sup> An inaccurate comment would be one that the respondent thought was true about the specific technology (e.g., TiVo is not a type of DVR) but was actually false.

interview technique used, so as to potentially glean more information from the rest of the participants. Given the small subject pool being used, this maximized the possibility of eliciting quality information from the participants.

The observations by the research team (without quantified data) led to the belief that questions of this degree of complexity served to confuse participants so that paraphrasing was not the correct technique. That is, participants were unable to repeat back the question because they were confused about what was being asked. Perhaps the verbiage used in our questions required a higher level thought process than is typical in survey research, which put an additional cognitive burden on the participants. As Ericsson and Simon (1993) have theorized, a heavy cognitive burden could cause the survey question to drop out of short-term memory. The end result was an undesirable, yet typical response of, “pretty much what you said.”

Another possibility is that people had difficulty paraphrasing our questions because it went against their pre-existing notion of what they thought the object was (Tourangeau, 1984). That is, the paraphrasing task forced participants to go against their schema of the object in question. Participants were processing in a top-down fashion, and thus were unable to fully comprehend our definition. Evidence for this comes from the confusions indicated by people during the interview. One common misconception was that On Demand was a movie only service. However, in reality, On Demand includes both TV programs and movies, which was stated in our description of the service. It could be that participants were unable to cognitively connect On Demand to television shows.

Of the first 11 participants, nine of them were confused about one of the three types of technology. Further, only two of them were able to clear up their confusions after extensive probing. This indicated that our survey questions likely needed modifications, but, more importantly, that we were using the wrong cognitive interview approach. If these conjectures were true, using an approach that led the participant to process in a more “piecemeal” fashion could yield higher quality interviews. Thus, in the second phase of our interviews, we tested this idea.

#### 4.2 Diary Design

For the diary design part of the cognitive interview, our instructions were very general. We simply asked participants to “think out loud” as they went through the diary. If questions were directed to the interviewer during this time, he or she deflected the question back to the participant with a response of, “However you would answer that if I wasn’t here.”

Our first goal for the design portion was to assess how participants would do on the scenarios we read them. Some found it easier than others to enter the programs in the scenarios. As to be expected, this normally depended on whether or not they had the types of technology in the scenarios and how well they remembered their diary-keeping experience.

Our attempt to learn about our second goal, which diary design was most effective, did not work as we intended. Instead, we found a recency effect for the presentation of the diaries. That is, whichever of the diary designs was presented last was the one participants liked the best. This was true for nine out of 11 of the first participants (82%). As stated earlier, there were eight total diary designs being tested, with each participant presented with two of the formats. Regardless of which two were presented, or in what order, the second design appeared to be the favorite.

One explanation for this could be that the participants saw the second diary as something new and different from the first. Since, for the majority of the participants, the scenarios were both challenging and mentally taxing, the idea of a new diary design might have had a contrastingly positive effect. This contrast effect could have potentially led to participants’ biased attitudes towards the second diary presented. Ostrom and Upshaw (1968) first theorized about contrast effects by postulating that, when first presented with a strongly positive or negative item, people tend to contrast items that follow. This leads to either a pronounced primacy or recency effect, depending on the valence of the extreme item. In the case of the diaries, it is possible that the difficulty participants had with the first diary led them to have an increased positive endorsement of the second diary, leading to a recency effect (see Schwarz and Hippler, 1991). If this reasoning is correct, then it should also be the case that presenting a third diary design would somewhat mitigate this effect. To this end, for the remaining interviews, a third design was introduced after each participant reacted to the second.

#### 5. Phase Two - Method

For the remainder of the cognitive interviews, the goal was still as it had been before: to learn as much as possible about the telephone recruitment questions and the diary test designs. It was agreed upon by all members of the research team that a revised approach to the interview should be created. Thus, instead of using the immediate retrospective technique, a concurrent approach was adapted. In this approach, participants were asked to think aloud as to how they would describe each of the types of the technology of interest. Interviewers probed when necessary to extract all knowledge the person had in regards to each

question. The probes used by interviewers started general (“Please describe what a DVR is.”) and became more tailored (“How well does this describe a DVR?”). After the participant had described the technology in detail, the phone recruitment question was read to them and they were asked to tell what they were thinking as they listened.

For the diary design part of the interview, we maintained the same technique as in Phase 1. The lone change to this part was to present a third diary design after reaction had been given to the second. It was our belief that comments about the third design might be less susceptible to order presentation effects. As a further test of this idea, the interviewer kept the third design out of sight of the participant until discussion about the first two designs had ceased. From here, the procedure mirrored that of the second diary presentation, where the participant was asked to think aloud in regards to the instructions or any other thoughts they had.

**6. Phase Two – Results**

In general, it was unanimous among members of the research team that our revised approach appeared to undoubtedly increase the amount of ideas suggested by the participants. They seemed more comfortable with the process and offered many more comments and descriptions as to what they felt each of the types of technologies were. Further, they listened more to the phone recruitment questions and suggested alternatives to the wording. Overall, our research team felt much better in regards to the second set of interviews.

However, how did the data from these interviews compare to the first set of data? To answer this, we again examined how many suggestions were offered by the participants in regards to each type of technology. As can be seen in Table 1, the increase in responses from our participants was overwhelming compared to the first phase of interviews. First, in regards to comments and suggestions about the four technology questions (i.e., Digital cable, On Demand, DVR standalone, and DVR with cable), participants gave an average of 7.05 responses (compared to only 5.0 previously). Secondly, additional probing revealed an average of 6.3 comments per participant, an increase of 1.7 from the first approach.

Moreover, even though we moved away from tailored probes in favor of more general probes, we received more consistent responses among our second group of participants. For example, exactly half of the second group identified On Demand as a “library of shows and movies”, a phrase we did not have in our

**Table 1: Comparison of responses in Phase 1 vs. Phase 2 (per respondent)**

	Comments about technology questions	Additional Comments
Phase 1 (n = 11)*	5.0	4.63
Phase 2 (n = 20)	7.05	6.30

\*There were 13 respondents interviews during this phase, but data is missing for two of them.

initial description for the term. This level of consistency enabled us to feel confident about modifying our description to include this term.

Overall, the data from the second phase, in which interviewers probed concurrently with the questions, led to much more elaboration from our participants. They responded more, gave more detailed answers, and seemed more comfortable with the think aloud process.

Results of presenting a third diary to our households indicated that there did appear to be a reduction in the recency effect by presenting a third diary design. First, as can be seen in Table 2, across all participants from both phases, 21 of 31 showed a preference for the second diary over the first, indicating an effect based on presentation order. The recency effect for the second phase was not quite as strong as the earlier phase, with 12 of 20 respondents indicating a preference for the second diary design. Of the remaining 40 percent, seven preferred the first design and one person was undecided.

Further examination of the 12 participants who showed a preference for the second design revealed that only half indicated liking the third design best of all (See Table 3). Thus, although the numbers are small, it does appear that introducing a third diary design mitigated the recency effect observed in the dichotomous comparison.

**7. Discussion**

The results of the comparison of these two phases tend to indicate that the cognitive interview approach used does matter in some instances. DeMaio and Landreth (2004) have recently tested the idea that different techniques may lead to different results. In their fine-grained analysis, they demonstrated that different techniques used by three separate research teams did lead to different results. They concluded that all three research teams were able to unearth problems in their questionnaires using different interviewing techniques. Further, although they used different types of probes, they were all able to get their respondents to think aloud.

**Table 2: Percentage of Respondents Who Chose First or Second Diary Design**

	<b>% Choosing First Design</b>	<b>% Choosing Second Design</b>	<b>Total %</b>
<b>Phase 1</b>	18.2	81.8	100.0
<b>Phase 2</b>	35.0	60.0	95.0*
<b>Total %</b>	29.0	68.0	97.0*

\*One participant from Phase 2 was undecided.

**Table 3: Percentage of Respondents Who Chose Third Design in Phase 2**

<b>Original Choice</b>	<b>% Choosing Original Choice</b>	<b>% Choosing Third Design</b>	<b>Total %</b>
<b>First</b>	50.0	50.0	100.0
<b>Second</b>	33.3	50.0	88.3*

\*Two participants were undecided between the second and third design.

After each team revised the original questionnaire based on the feedback from cognitive interviews, the questionnaires were analyzed for problems. Once again, the results were inconclusive, although it did appear that two of the teams outperformed the other.

The current research looked at how different cognitive interview techniques within the same research team differed. While our original intent was not to compare these methods, this, nonetheless, was a direct test of the interplay among the two most common techniques. By using the same interviewers and the testing the same questions, we were able to examine differences in participant response, as opposed to differences in interviewer characteristics.

We initially posited that perhaps the complexity of the question combined with asking the participants to recall the exact wording placed a cognitive burden on them. For the second approach, in which participants were first given the opportunity to define the terms before the interviewer read the question, it appears that there was more room for a person's construction of the technology in question. This fits nicely into the concept of bottom-up processing (see Tourangeau, 1984). In instances where a person only has a vague idea or concept, the mind creates a mental image of what is known. In our case, participants could have used thinking out loud to construct a working definition of what, say, digital cable is in order to better understand it. This is a stark contrast from the first approach, in which the interviewer read a description of the vague technology in question. It is not that the participants did not know anything about

the technology, just that they had difficulty putting into words what it was. By enabling the participants to have the first crack at it, it is possible that this enhanced their processing ability.

The implications of this suggest that, maybe, at times, some cognitive interviewing approaches are better than others. In this case, perhaps the use of complex technical terms was better served using a concurrent approach as opposed to retrospective. Then, participants can say, "Yes, I have digital cable and I know this because I get so many channels and I have access to a lot of features." Thus, the link between a person's ability to think out loud and their cognitive processes could have something to do with the complexity of the task and the capacity of one's short-term memory.

For the second part of these cognitive interviews, we presented different diary designs to participants and asked them to think aloud about which designs they liked best as a tool for entering in TV viewing data. Initially, there appeared to be a presentation order effect in which, regardless of which designs were presented, the second of two diaries was more well-received. We attempted to overcome this bias by introducing a third diary design after the second. Our results showed that there was some evidence for a tempering of this effect, as the preferences seemed to be based more on the diary than on the order in which it was presented.

In general, most cognitive interviews are a means of pre-testing survey questions. The effects of using this method for pre-testing actual designs and graphics is unknown and needs further investigation to show evidence for its validity.

## 8. Conclusion

While the positions described here provide grounds for a theoretical underpinning to cognitive interviewing approaches, the evidence is not strong enough to make a general claim. First, while this is a good example of how various cognitive approaches can affect a participant's responses, it was not designed to be a controlled experimental test. The initial goal was to pre-test telephone recruitment calls dealing with new technology that directs which Nielsen Diary will be sent to the household. When we changed our method in hopes of improving results, the results between the two techniques were markedly different—the second provided better responses and quality suggestions for changes to our questionnaire. This, then, begged the question of why. What cognitive structures caused such a large difference in responses from one approach to another?

To this end, the initiatives put forth in this paper were all posited as an answer to the above question.

Future research could be designed to test these concepts more stringently and directly. One possibility would be to have the same research team conduct a series of cognitive interviews, some using a strict retrospective, tailored type approach, others using a more general, concurrent approach. Using a coding approach as detailed by DeMaio and Landreth (2004; also see Rothgeb, Willis, and Forsyth, 2001) could enhance the conclusions drawn here in order to see exactly what types of comments are being made in regards to the questionnaire (i.e., comprehension, retrieval, etc.). Further, a closer look could be taken at exactly what types of retrospective and concurrent probes are asked during these interviews (i.e., paraphrasing, think-aloud, tailored follow-ups, etc.). Thus, this research is only a precursor for other potential studies that could examine these issues.

### References

- Daugherty, S., Harris-Kojetin, L., Squire, C., & Jael, E. (2001). Maximizing the quality of cognitive interviewing data: An exploration of three approaches and their informational contributions. *Proceedings of the Annual Meeting of the American Statistical Association*. Retrieved March 2006, from American Statistical Association Website: <http://www.amstat.org/sections/srms/Proceedings/y2001/y2001.html>.
- DeMaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215-251.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*, 2<sup>nd</sup> ed. Cambridge, MA: MIT Press.
- Forsyth, B. H., & Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.), *Measurement errors in surveys*. New York: Wiley.
- Hess and Singer (1995). The role of respondent debriefing questions in questionnaire development. Retrieved March 2006, from American Statistical Association Website: [http://www.amstat.org/Sections/Srms/Proceedings/papers/1995\\_187.pdf](http://www.amstat.org/Sections/Srms/Proceedings/papers/1995_187.pdf)
- Martin, E. (2004). Vignettes and respondent debriefing for questionnaire design and evaluation. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley.
- Nisbett, R., & Wilson, T. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- Ostrom, T. M., & Upshaw, H. S. (1968). Psychological perspectives and attitude change. In A. G. Greenwald, T. C. Brock, and T. M. Ostrom (Eds.), *Psychological foundations of attitudes*. New York: Academic Press.
- Redline, C., Smiley, R., Lee, M., DeMaio, T., & Dillman, D. (1998). Beyond concurrent interviews: An evaluation of cognitive interviewing techniques for self-administered questionnaires. Retrieved March 28, 2006, from US Census Bureau Website: <http://www.census.gov/srd/papers/pdf/sm98-06.pdf>.
- Rothgeb, J., Willis, G., and Forsyth, B. (2001). Questionnaire pretesting methods: Do different techniques and different organizations produce similar results? *Proceedings of the Annual Meeting of the American Statistical Association*. Retrieved March 2006, from American Statistical Association Website: <http://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00476.pdf>.
- Schechter, S., & Beatty, P. (1994). Conducting cognitive laboratory tests by telephone. *National Center for Health Statistics, cognitive methods staff, working paper series, no. 8*.
- Schwarz, N., & Hippler, H. J. (1991). Response alternatives: The impact of their choice and ordering. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.), *Measurement errors in surveys*. New York: Wiley.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.
- Willis, G. B. (2001). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.