

BEYOND CONCURRENT INTERVIEWS: AN EVALUATION OF COGNITIVE INTERVIEWING TECHNIQUES FOR SELF-ADMINISTERED QUESTIONNAIRES

Cleo Redline, Richard Smiley, Meredith Lee, and Theresa DeMaio, U.S. Bureau of the Census, and Don Dillman, Washington State University
Cleo Redline, Bureau of the Census, Room 3125--4, Washington, D.C. 20233

Key Words: Questionnaire Design, Cognitive Interviews.

1. INTRODUCTION

In the last decade, the application of cognitive psychology to the development of survey questions and questionnaires has advanced greatly. Evidence to this effect comes from the fact that many institutions from around the world have come to rely considerably upon laboratories employing cognitive interviewing techniques to evaluate survey questions and questionnaires (see, e.g., DeMaio and Rothgeb, 1996; Akkerboom and Dehue, 1997). Two of the most widely described cognitive interviewing techniques are the *concurrent* and *retrospective* interviews (see, e.g., Ericsson and Simon, 1980; Forsyth and Lessler, 1991). In concurrent interviews, subjects are asked to verbalize the process they go through to answer a question as they progress through a questionnaire. In retrospective interviews, subjects are asked at the end of the interview to verbalize thoughts about questions they answered earlier in the questionnaire.

Survey questions and questionnaires can be broadly divided into two types: interviewer-administered or self-administered. The cognitive technique commonly employed to evaluate interviewer-administered questionnaires is the concurrent interview (see, e.g., DeMaio et al., 1993). This seems a reasonable choice, given that interviewer-administered questionnaires are communicated verbally and Ericsson and Simon (1980) asserted that the concurrent technique will provide the most reliable source of data concerning subjects' thought processes when the task requires subjects to verbalize information attended to in short-term memory as verbal code. However, Ericsson and Simon also argue that the concurrent technique is not advisable under certain conditions. One example they give is when a task requires subjects to verbalize visual information. They suggest that visual encodings will require a verbal recoding, and this has been shown to affect the way the information is processed.

Jenkins and Dillman (1997) proposed a model of self-administered questionnaire design that acknowledged the importance of visual language in addition to verbal language. The model was comprised of two major

components: (1) subject matter understanding¹ and (2) navigation. The term 'subject matter understanding' broadly refers to the way in which the questionnaire's verbal and visual language is arranged to convey the meaning of each intended question. The term 'navigation' is used to denote the way in which the questionnaire's visual and verbal information gives rise to its prescribed path.

Thinking in terms of this model led us to deduce that self-administered questionnaires may contain features, such as the wording of questions and response categories, that lend themselves to being evaluated through the concurrent technique and features, such as the questionnaire's visual landscape, that do not. As a result, we hypothesized that both the concurrent and retrospective techniques would provide different, but valuable information about respondents' form filling behavior, and that both should be used to adequately evaluate a self-administered questionnaire. Specifically, we hypothesized that:

1. the concurrent interview will provide more information concerning respondents' understanding of the survey's subject matter than the retrospective interview,
2. the concurrent interview will provide information that is less representative of a respondent's natural navigation in questionnaires than the retrospective interview.
- 3.

An opportunity to test these hypotheses arose in 1995 when we were asked to conduct cognitive interviews with three newly designed self-administered forms for possible use in Census 2000. We designed the study to include a small-scale experimental comparison of the two interviewing methods. This paper presents the results of that comparison. Due to time and resource constraints, the analyses and results presented in this paper are limited to the second hypothesis. We conclude with a discussion of the results and their implications for the future.

2. CONCEPTUAL BACKGROUND

In concurrent interviews, respondents are asked to read

¹Jenkins and Dillman (1997) originally used the term 'information organization.'

and think aloud and to respond to two sets of questions, those offered visually in the questionnaire materials and those offered orally by the interviewer. Potentially this may affect respondents' behavior in a number of ways. Respondents may read/process less because their attention must be divided between the questionnaire materials, their need to think aloud, and their need to listen to the interviewer (Biehal and Chakravarti, 1989). Or conversely, respondents may read/process more in the concurrent interview because their attention is artificially inflated by the instruction "to read and think aloud" in the presence of an interviewer. However, another possibility is that the concurrent interview may affect what respondents choose to focus their attention on. For instance, one might argue that respondents are encouraged to focus their attention on the subject-matter content in a concurrent interview.

2.1 Subject-Matter Understanding

Preliminary research suggesting that respondents do read/process or focus more on the subject-matter content in concurrent interviews than in a more realistic self-administered setting comes from Gerber et al. (1997). The researchers found that respondents in a classroom experiment condition had somewhat less success listing vignette characters correctly than respondents in a concurrent interview condition. They concluded that the concurrent interview may lead respondents to pay somewhat more attention to the questionnaire's content than a self-administered setting.

2.2. Navigational Qualities

An area in which some researchers have argued the opposite--that the concurrent interview causes respondents to read less--is with skip instructions. Concurrent interviews with self-administered questionnaires have revealed that respondents commonly overlook skip instructions (see, e.g., Gower, 1989). Critics of the concurrent method argue that respondents might overlook the skip instructions because the interviewer distracts the respondent (see, e.g., Schechter et al., 1996). However, a competing hypothesis first offered by Jenkins and Ciochetto (1993) is that respondents may overlook the skip instructions because of form effects, that is, because the skip instructions are located to the right of where respondents' eyes are naturally traveling.

Experimental evidence suggests that critics of the concurrent method can only be partially correct at best. Turner et al. (1992) report that faulty execution of skip instructions in self-administered questionnaires occurred to varying degrees, depending on the skip instruction's format and in the absence of any interviewer at all. This strongly suggests that the overlooking of skip instructions must be at least partially due to format considerations. However, it leaves open the question of whether the

faulty execution of skip instructions is exacerbated in a concurrent interview.

3. METHODOLOGY

In 1995, a small-scale experiment comparing the two interview methods was included as part of a larger study whose main objective was to evaluate three newly designed census short forms, known as the green booklet, yellow booklet, and yellow oblong, respectively (see Dillman et al., 1996 for the larger study). Respondents completed the questionnaires in one of two orders: (1) either the green booklet, yellow booklet, and yellow oblong or (2) the yellow oblong, yellow booklet and green booklet.

The two questionnaire orders and two cognitive interviewing treatments were randomized and preassigned across 55 interviews such that:

- 13 interviews were conducted concurrently and received the green booklet first;
- 15 interviews were conducted retrospectively and received the green booklet first;
- 15 interviews were conducted concurrently and received the yellow oblong first;
- 12 interviews were conducted concurrently and received the yellow oblong first.

Twenty of the 55 volunteers were interviewed at the Census Bureau; the remaining were conducted at Washington State. An attempt was made to video- and audiotape all of the interviews at both institutions. In three cases where these videotapes failed, however, the interviews were audio taped.

3.1 Respondent Behavior Codes

Dillman et al. (1996) developed a set of respondent behavior codes. At the Census Bureau, the interviewers personally coded their own interviews from the videotapes (or from the audiotapes) shortly after the interview was conducted (a week or two at most). At Washington State, coders coded the interviews shortly after the interviews as well. Approximately 20 behavioral codes per questionnaire were applied to describe respondents' handling of the questionnaire. A subset of these described navigational errors respondents made. For example, respondents were expected to start at the front of the form, and looking at the back of the form first was considered an error. Conversely, never looking at the front page was considered an error, as was starting to answer questions on the right hand side of a page rather than the left. Whether respondents mis-executed the skip instruction on the yellow oblong form was based on a review of respondents' completed forms, as opposed to the previous codes which were based on observing respondents as they completed the forms.

As noted above, respondents filled out all three of the questionnaires in one of two orders. We were concerned that learning may have occurred as respondents filled out

multiple forms. Preliminary analysis determined that order effects were in fact observed on some of the navigational codes, so we limited our analyses to only those codes applicable to the first forms completed by the respondents: either the green booklet or yellow oblong.

3.2 Item Non-Response Data

We defined item non-response as the percent of persons reported on the form for whom a response was required and for whom none was obtained. Data were keyed from the three questionnaires for item non-response analyses. First, a binary code ('0' for a non-response and '1' for a response) was keyed for every conceivable answer space on the questionnaire, including each and every check box and write-in answer spaces. The data were collapsed as follows: for the questions which exclusively contained check boxes, if respondents marked a response in one or more of the check boxes, the item was coded with a '1,' meaning a response was present. For the questions that had check boxes followed by write-in boxes, if respondents provided a response in either the check boxes or the write-in spaces, the item was considered answered. Questionnaire items were coded for persons 1-5. Data gathered from the continuation roster were not examined.

For the purposes of this paper, the 11 items which were the same across persons within a questionnaire and which were the same across questionnaires were examined: last name, middle initial, first name, sex, month, day, year of birth, age, Hispanic, race and another residence. Again, the analyses were limited to the first questionnaire administered in the series.

4. RESULTS

4.1 Respondent Behavior Codes

In this paper, we hypothesize that the concurrent interview will provide information that is less representative of respondents' natural navigation through questionnaires than the retrospective interview. To investigate this hypothesis, we examined the navigational errors for which codes were developed. As Table 1 shows, there was very little difference between the concurrent and retrospective methods in the majority of navigational error rates nor the weighted error rate across all of the codes. Of the 12 navigational indicators, only one showed a significant difference. This was the error of going down a "postcard" form as if it were a columnar form. The green booklet had a "postcard" design in which respondents were supposed to stop reading halfway down the page and return to the top. So here, going down the page in a columnar fashion was an error. Over 45 percent of the respondents in the concurrent interviews erroneously filled out the green booklet in a columnar fashion, while less than 10 percent of those in the retrospective interviews did so.

Since previous research has shown that literacy level can affect a respondent's ability to understand and

complete documents (see e.g., Kirsch et al., 1993; Gerber and Wellens, 1995), we examined the education level of respondents by interview method. We used two levels of education defined as follows: the low educated group consists of respondents who have a high school education (including a GED) or less; the high education level consists of respondents who have at least some college education.

Table 2 presents the results of this analysis. In terms of the green booklet, respondents in the retrospective method were significantly more educated than respondents in the concurrent method. In terms of the yellow oblong, the table shows that while there were some differences in the education level of respondents in the concurrent and retrospective conditions, these differences were not significant. This suggests that despite our efforts to randomize treatment conditions, the small treatment sizes (of from 10 to 15) resulted in unbalanced treatment groups.

Table 3 shows that the observed difference between the concurrent and retrospective treatments is entirely concentrated among the less educated respondents. Almost 55 percent of the less educated respondents in the concurrent treatment filled out the form incorrectly, while none of the less educated respondents in the retrospective treatment made this error.

It should be noted, however, that navigational codes which contained any amount of reading behavior (i.e., verbal language) worked relatively well in the concurrent setting, but not nearly as well in the retrospective. For instance, determining whether respondents read the skip instructions was based on a sample of 14 in the concurrent, but dropped to a sample of 4 in the retrospective.

4.2 Item Non-response

Another indicator of respondents' navigational behavior is whether or not respondents provided answers to the items. Table 4 shows that overall, item non-response did not differ by interview method.

5. DISCUSSION

It does not appear to be the case in our study that, overall, the retrospective interview provided different information than the concurrent interview. This is good news because it suggests that concurrent interviewers had less effect on respondents' navigational behaviors than originally supposed. It also provides more evidence in support of the fact that respondents overlooked the skip instructions in the concurrent interviews as a result of form effects, not interviewer effects.

We did, however, find one instance in which the results provided by the two methods differed, and it is interesting to speculate why this particular behavior was susceptible to method effects when the others were not. We discovered that the less educated respondents were more

likely to incorrectly continue reading down the page in the concurrent interview than the retrospective. Many may assume that the less educated respondents were more easily distracted by the interviewer, but we do not subscribe to this, since these same respondents were not distracted by their interactions with the interviewers when it came to any of the other navigation behaviors. Upon closer inspection, it turns out that the columnar error was the only one which appeared to rely entirely on visual cues (that of seeing two green person "postcards" outlined in black and separated by a white space), whereas the other navigational codes appeared to have some amount of verbal language associated with them. It seems plausible that the less educated respondents were so intently focused on the questionnaire's verbal content in the concurrent interview that they did not perceive the purely visual language because they were reading aloud in the presence of an interviewer.

We now have two studies--ours and the Gerber et al. (1997) study--which have demonstrated differences between a verbal versus a silent condition. Although the difference we found was slight (that is, one navigational code out of twelve), it is possible that the reason behind the differences is the same. The Gerber et al. found greater understanding of the questionnaires' subject matter content (which one can categorize as largely verbal) in concurrent interviews and we found that respondents may have focused on the verbal content as well, perhaps to the detriment of the visual. We need to determine if there is anything to this. If there is, maybe our original definition of navigation needs to be refined (such that navigation which relies on verbal language is distinguished from that which relies on purely visual cues).

6. CONCLUSION

In this paper we have compared the results of the concurrent interview with the retrospective interview, with an eye toward determining if they provide different kinds of information. We are in the exploratory stage of this work, in terms of determining what the best procedures are for testing multiple self-administered forms, eliciting information about the navigational aspects of these forms, and setting up experimental evaluations of the methods. Thus, we view our results as preliminary food for thought.

Before more definitive conclusions can be drawn, this research needs to be replicated with a larger sample, and because it was hard to determine if respondents had read information in the retrospective interview, eye-movement studies are necessary to improve our understanding of what respondents read in retrospective interviews. Then, these ideas need to be tested with different questionnaires. A great deal of research has been conducted with the decennial short forms. As a result,

one would hope that the navigational features of this questionnaire are clearer than the navigational features of less studied questionnaires. An ideal test would be to design a questionnaire with known subject matter understanding and navigational problems. Only after these kinds of further studies are conducted will we know with any certainty whether or not the two methods provide different kinds of information.

ACKNOWLEDGMENTS

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. We would like to thank Elizabeth Martin and Beth Nicolas for reviewing an earlier draft of this paper.

REFERENCES

- Akkerboom, H. and Dehue, F. 1997. "The Dutch Model of Data Collection Development for Official Surveys." *International Journal of Public Opinion Research*. Vol. 9, No. 2, pp. 126-145.
- Biehal, G. and Charkravarti, D. 1989. "The Effects of Concurrent Verbalization on Choice Processing." *Journal of Marketing Research*. Vol 26: 84-96.
- DeMaio, T.J. and Rothgeb, J.M. 1996. "Cognitive Interviewing Techniques: in the Lab and in the Field," in N. Schwarz and S. Sudman (eds.), *Answering Questions*. California: Jossey-Bass Inc., pp. 177-198.
- DeMaio, T., Mathiowetz, N., Rothgeb, J., Beach, M.E., and Durant, S. 1993. *Protocol For Pretesting Demographic Surveys at the Census Bureau*. Report of the Pretesting Committee, U.S. Bureau of the Census, Washington, DC.
- Dillman, D.A., Jenkins, C.R., Martin, B., and DeMaio, T. 1996. "Cognitive and Motivational Properties of Three Proposed Decennial Census Forms." Center for Survey Methods Report, U.S. Bureau of the Census, Washington, D.C.
- Ericsson, K.A. and Simon, H.A. 1980. "Verbal Reports as Data." *Psychological Review*, Vol. 87: 215-251.
- Forsyth, B.H. and Lessler, J.T. 1991. "Cognitive Laboratory Methods: A Taxonomy" in P.B. Biemer, R.M. Groves, L.E. Lysberg, N.A. Mathiowetz, and S. Sudman, (eds), *Measurement Errors in Surveys*. New York: John Wiley and Sons, pp. 393-418.
- Gerber, E., Keeley, C., and Wellens, T. 1997. "Census Rules and Rostering Decisions: A Vignette Study." *Proceedings of the Section on Survey Methods Research*. American Statistical Association.
- Gerber, E. and Wellens, T. 1995. "Literacy and the Self-Administered Form in Special Populations: A Primer." *Proceedings of the Section on Survey Methods Research*. American Statistical Association.
- Gower, A.R. and Dibbs, R. 1989. "Cognitive

Research: Designing a 'Respondent Friendly' Questionnaire for the 1991 Census." *Proceedings of the Bureau of the Census Fifth Annual Research Conference*. pp. 257-266.

Jenkins, C. and Ciochetto, S. 1993. "Results of Cognitive Research on the Multiplicity Question from the 1991 Schools and Staffing Survey Student Records Questionnaire." Center for Survey Methods Report, U.S. Bureau of the Census, Washington, D.C.

Jenkins, C.R. and Dillman, D.A., 1997. "Towards a Theory of Self-Administered Questionnaire Design," in L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, and D. Trewing (eds.), *Survey Measurement and Process Quality*. New York: Wiley-Interscience.

Kirsch, I. S, Jungeblut A., Jenkins, L., and Kolstad, A. 1993. *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*. Educational Testing Service under contract with the

National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education.

Schechter, S., Blair, J., Hey, J.V. 1996. "Conducting Cognitive Interviews To Test Self-Administered and Telephone Surveys: Which Method Should We Use?" *Proceedings of the Section on Survey Methods Research, American Statistical Association*. American Statistical Association.

Turner, C.F., Lessler, J.T., George, B.J., Hubbard, M.L. and Witt, M.B. 1992. "Effects of Mode of Administration and Wording on Data Quality" in C.F. Turner, J.T. Lessler, and J.C. Gfroerer, (eds), *Survey Measurement of Drug Use Methodological Studies*, Washington, DC: National Institute of Drug Abuse, U.S. Department of Health and Human Services, pp. 221-243.

Table 1. Percentage of navigational errors by interview method and form type.

Navigational Codes	Concurrent (%)	N	Retrospective (%)	N	Fisher's Exact p-value
Green Booklet Form					
Started by looking at back page <u>first</u>	0.0	12	7.1	14	n.s.
Did <u>not</u> ever look at front page	7.7	13	0.0	15	n.s.
Filled out person spaces in columnar fashion	46.2	13	7.7	13	0.037
Did <u>not</u> read "note" in last person space reported	36.4	11	42.9	7	n.s.
Did <u>not</u> put form back in envelope correctly	30.8	13	21.4	14	n.s.
Yellow Oblong Form					
Started by looking at back page <u>first</u>	53.3	15	54.6	11	n.s.
Did <u>not</u> ever look at front page	33.3	15	8.3	12	n.s.
Started under any of the person triangles	0.0	13	8.3	12	n.s.
Did <u>not</u> read "Go to Next Person" inst. in P1 box	42.9	14	25.0	4	n.s.
Trouble Understanding "Go to Next ..." instr.	63.6	11	60.0	10	n.s.
Mis-executed skip instruction	0.0	14	10.0	10	n.s.
Did <u>not</u> put form back in envelope correctly	35.7	14	10.0	10	n.s.
Weighted average across all items	34.8	158	32.6	132	n.s.

Table 2. Percentage of less educated respondents by interview method and form type.

Level of Education	Green Booklet			Yellow Oblong		Fisher's Exact p-value
	Concurrent (%)	Retrospective (%)	Fischer's Exact p-value	Concurrent (%)	Retrospective (%)	
High School or Less / GED	75.0	33.3	0.054	64.3	50.0	n.s.
N	12	15		14	10	

Table 3. Percentage of respondents who erroneously filled out the green booklet person spaces in columnar fashion.

Level of Education	Concurrent (%)	N	Retrospective (%)	N	Fisher's Exact p-value
High School or Less / GED	55.6	9	0.0	5	0.063
Some College	25.0	4	12.5	8	n.s.

Table 4. Percent item non-response by form type and interview method.

Question Item	Green Booklet			Yellow Oblong		
	Concurrent (%)	Retrospective (%)	Chi-sq. p-value	Concurrent (%)	Retrospective (%)	Chi-sq. p-value
Wt. avg. across 11 items	8.1	8.2	n.s.	10.0	9.7	n.s.
N	430	500		510	300	