

# EFFECTS OF QUESTIONNAIRE DESIGN ON THE QUALITY OF SURVEY DATA

MARIA ELENA SANCHEZ

**Abstract** The importance of format, graphic layout, and question routing instructions on the quality of survey data has been acknowledged for self-administered survey instruments, but the significance of these factors for questionnaires administered by interviewers has not been recognized equally. This paper examines two studies in which interviewers used different questionnaire designs to administer the same set of survey questions to randomly selected adult respondents in housing units that made up area probability samples of the Detroit metropolitan area. The paper presents empirical evidence of specific effects (questions skipped in error, unprobed answers) directly attributable to the choice of questionnaire design. The analysis shows that questionnaire design choices can either help or hurt the quality of data collected by interviewers. Furthermore, the behaviors of experienced and inexperienced interviewers are affected in similar ways. In other words, interviewing experience does not compensate for format deficits in the design of survey instruments.

Questionnaire design is frequently overlooked as an important aspect of the development of field instruments and as a potential source of independent effects on survey estimates. Discussions about topics such as question formatting options, graphic layout, integration of interviewer recording tasks for complex question series, and optimal routing strategies are frequently absent from published works that otherwise deal with questionnaire development.

The quality of questionnaire design is generally recognized as an important factor for self-administered instruments (Dillman 1978, 1983); however, the importance and special problems of design for questionnaires to be administered by interviewers has not been ac-

MARIA ELENA SANCHEZ is a survey methodologist at the Agency for Health Care Policy and Research. I wish to thank the following: Duane F. Alwin, who generously granted permission to use the data; Merilynn R. Dielman, especially, and Dawn R. Von Thurn, who facilitated access to study tapes and questionnaires on more than one occasion; and an unidentified reviewer who offered very helpful comments and suggestions on an earlier draft of this paper.

knowledgeed equally. For example, in the chapter on questionnaire construction, Babbie (1979) deals almost exclusively with the format of self-administered instruments, which, of necessity, are less complex and burdensome than the average standardized instrument handled by survey interviewers.

Sudman and Bradburn (1983) discuss interviewer instruments and self-administered instruments in their chapter on questionnaire format, which offers valuable practical advice. But only one of several extant formatting styles for interviewer instruments is reviewed in any detail, and the format choices these authors recommend stem more from common sense and experience than from systematic evidence, as they are the first to admit. Nonetheless, Sudman and Bradburn's presentation is more extensive and rigorous than the typical discussion of elementary design issues found elsewhere, even in otherwise very comprehensive survey texts (Moser and Kalton 1974). Sheatsley (1983) is somewhat of an exception in that he attempts to provide explicit rules about dos and don'ts in the area of questionnaire design. Even so, this author devotes approximately two and one-half pages to questionnaire formatting issues in a chapter about questionnaire construction that runs some 40 pages in length.

All of these authors agree that inadequate questionnaire design will increase measurement error, among other things. But the literature offers scant empirical evidence to support this claim (although few survey practitioners would dispute the point based on anecdotal evidence), and consequently, we lack estimates of the magnitude of effects attributable to questionnaire design.

In this paper we examine two studies that used different questionnaire designs for the same set of survey questions and present evidence of effects directly attributable to the choice of design. We also evaluate the importance of interviewer experience in accounting for these effects.

### **Background Information about the Studies Compared**

In the spring of 1987, the department of sociology at the University of Michigan conducted its annual Detroit Area Studies (DAS) survey as part of a course designed to teach students about survey research. The survey topic was religion in Detroit. An area probability sample design was used to identify eligible households in the Detroit tri-county area. Face-to-face interviews were conducted in each household with one randomly chosen resident aged 21 years or older.

A few months later the DAS principal investigator secured additional funding to enlarge the total study sample size and, in the fall of 1987,

the DAS survey was replicated in the same Detroit counties by the Field Section of the Survey Research Center (SRC), University of Michigan. The sample size and design for the replicate study, Religion in Detroit (RID), were equal to those used in the DAS study. The sample design for both studies was controlled by SRC's Sampling Section.

The DAS study completed 554 interviews and achieved a 70 percent response rate. Twenty-one DAS students were responsible for 32 percent of the interviews, as a course requirement. The remainder of the work was handled by 29 experienced SRC field interviewers under the direction of DAS personnel.

The RID study completed 494 interviews, for a response rate of 68 percent. Seven newly trained SRC interviewers with no prior field interviewing experience completed 26 percent of the interviews. The remaining interviews were the responsibility of 19 experienced SRC field interviewers. Fourteen of these interviewers had also worked on the DAS study.

The field procedures for both studies were comparable and the contents of the survey instruments were identical. However, the format and graphic layout for a few DAS questionnaire items were changed in RID due to potential design problems that were identified in reviewing the DAS questionnaire.<sup>1</sup>

The average load of an RID interviewer was almost twice as large as the average load of a DAS interviewer (19 versus 11 completed interviews). The difference is largely due to the reduced work load of students. On the other hand, experienced and newly trained RID interviewers handled average interview loads of similar size (19 and 18 interviews, respectively), as did SRC interviewers who worked on both surveys (17 interviews in DAS versus 19 in RID). The median

1. The DAS and RID studies were not part of a controlled experimental design, but we believe our results are valid in spite of the design shortcoming: (a) The population and the sample design were identical for both studies. The mode of administration, the order, and the wording of questions in the two surveys remained the same, with changes restricted to differences in the layout and formatting of text on the printed questionnaire page. (b) Neither interviewers nor field supervisors were told that we had plans to compare the two studies, and nothing was therefore said about the kind of results we expected to find. The decision to produce new field materials for the RID study was not an unusual one. All studies entrusted to SRC's Field Section have materials reworked to ensure compliance with a variety of conventions and standards that apply to questionnaires and other field documents. (c) Both studies employed a mix of experienced and inexperienced interviewers, and the proportion of interviews completed by each of the two interviewer groups was roughly the same in the two studies. New RID recruits may have received more complete training in interviewing techniques than their counterpart student group, but students had a better understanding of the study objectives, having spent the better part of a semester discussing the study. Students also had participated in two informal pretests prior to the launching of the DAS study.

and mean values for each of these distributions are quite comparable in magnitude; thus, survey results are not unduly influenced by a few interviewers with unusually high numbers of completed interviews.

### The Original and Revised Questionnaire Designs

The questionnaire designs to be compared include (a) the placement in a grid of two questions dealing with the past and current religious affiliation of the respondent's children and (b) the format used in questions that required interviewers to probe for the name of specific Protestant denominations. Each gives rise to unique effects discussed in detail below.

#### QUESTIONS SKIPPED IN ERROR

The DAS study used the grid reproduced as figure 1 to collect information about the respondent's children. We are interested in effects associated with two of the questions: G11, which asks about the child's religious affiliation while growing up, and G12, which establishes the current religious affiliation for children aged 15 years or older.

Two features of the DAS design for questions G11 and G12 were identified as likely to make interviewers skip one or the other of these two questions in error:

1. The first is the stacked, as opposed to linear, arrangement of G11 and G12 in the DAS grid. After listing the children's names, interviewers filled the DAS grid from left to right for questions G8–G10, but from top to bottom at G11/G12. As a result, G11 answers in DAS could be entered by accident in the line reserved for G12 (leaving G11 blank and unasked); or the interviewer could overlook G12 altogether, moving on to ask about the next child after recording the answer for G11 in the correct line of the grid. These errors, due to the fact that the interviewer failed to ask the religious preference question, yield responses coded "Not Ascertained" (NA) at G11 or G12 in DAS.
2. Question G12 in the DAS study was additionally burdened with a contingency clause printed in the page of questions opposite the grid. Question G12 applied to children born in 1972 or earlier, and this meant the interviewer had to check G9 ("Year of Birth") before deciding whether to ask G12 or skip that line on the grid. In our analysis we look at the times this question was

04. INTERVIEWER CHECKPOINT: CHILDREN

(SEE PREVIOUS PAGE, G3 AND G4)

1. R IS RAISING OR HAS RAISED CHILDREN

2. R HAS NO CHILDREN BUT PLANS TO HAVE CHILDREN --> TURN TO PAGE 09, G03

3. R HAS NO CHILDREN, AND DOES NOT PLAN TO HAVE CHILDREN OR IS UNCERTAIN --> TURN TO PAGE 08, G01

G7. How I would like to get some more detailed information about the child(ren) you (have raised/are raising). Could you please tell me their first name, starting with the obsest? [ENTER NAMES IN GRID. REPORT G8-G11 FOR EACH CHILD.]

G8. Is the (1st, 2nd, etc.) a boy or girl?

G9. In what year was (he/she) born?

G10. Is (he/she) living with you?

G11. In what religion (was raised/is being raised)?  
Is it Protestant, Catholic, Jewish, some other religion, or none?  
(IF PROTESTANT): that specific denomination is that?

G12. (NONE IS 1973 OR BEFORE): that is (his/her) religious preference now?  
Is it Protestant, Catholic, Jewish, some other religion, or none?  
(IF PROTESTANT): that specific denomination is that?

G7. CHILD'S NAME (FIRST WITH SURNAME)	G4. SEX	G9. YEAR OF BIRTH	G11. RELIGION RAISED IN			G12. OTHER (SPECIFY)
			PROTESTANT DENOMINATION	CATHOLIC	JEWISH	
1ST			G11.			
			G12.			
2ND			G11.			
			G12.			
3RD			G11.			
			G12.			
4TH			G11.			
			G12.			
5TH			G11.			
			G12.			
6TH			G11.			
			G12.			
7TH			G11.			
			G12.			

— NONE TRAIL SEVEN CHILDREN: USE EXTRA CHILD SHEET.

Figure 1. The DAS questionnaire design for questions G11 and G12

skipped in error for children aged 15 years or older in DAS (coded "NA" at G12).<sup>2</sup>

In the revised RID design (fig. 2), the children's names are entered across the top of facing pages, and the interviewer administers the questions for each child's column, always reading and recording answers from top to bottom, as in any other questionnaire page. A checkpoint (RID item G12) was added to handle explicitly the contingency clause dealing with the child's age, and DAS item G12 was renumbered to G13 in the RID questionnaire.

The results reveal that, as expected, the DAS layout version yields a proportionately higher amount of "Not Ascertained" (NA) answers. In DAS, 3.47 percent of responses to G11 were NA, while in RID, only 0.54 percent of answers to G11 were NA. Similarly, for G12, 8.80 percent of responses were NA in DAS, but only 2.04 percent of responses to the corresponding question in RID (G13) were NA.

#### UNPROBED ANSWERS

The DAS survey and the RID study used the same battery of questions to collect information about the respondent's good friends (E49–E53) and siblings (H23–H27). The DAS design for the "good friends" series is reproduced as figure 3A; the RID version appears as figure 3B. In the battery of questions, religious affiliation is obtained at E53 for friends and at H27 for siblings. "Protestant" answers to E53 and H27 were to be probed at all times in order to obtain the names of specific Protestant denominations.

Close examination of figure 3A reveals the following about the manner in which interviewers were expected to handle the answers to E53 (and H27) in DAS. For "Catholic," "Jewish," and "None" answers to E53 (or H27), all the interviewer had to do was enter an *X* in the appropriate box of the grid. For "Other" answers, the grid instructed the interviewers to write ("Specify") the name of that religious preference inside the box. "Protestant" answers were to be handled in yet a different manner. Interviewers were supposed to probe the initial answer to the religious preference question with "What specific denomination is that?" in order to elicit a denomination name to be written inside the "Protestant Denomination" box of the grid.

2. Problems arising from asking G12 in error (when the age of the child did not warrant this) cannot be identified in the coded data, where the field has been forced to inappropriate (INAP), and are therefore not considered here. In addition, both studies collected information about deceased children, but DAS coded these entries "INAP" (000) in G12 while RID used "NA" (999) in G13. The RID entries for deceased children were identified through questionnaire lookups and recoded to "INAP" (000) for this analysis.

06. OMITTED

07. How I would like to get some more detailed information about the child(ren) you have raised/are raising). Could you please give me their first name(s), (starting with the surname)? (SECOND NAME IS GID). IF MORE THAN 10 CHILDREN, CONTINUE WITH EXTRA CHILD FIRST.)

	a. NAME 01	b. NAME 02
07. CHILD'S FIRST NAME (GIDIS 070)	1. MALE 2. FEMALE	1. MALE 2. FEMALE
08. Is (name 01, 02, etc.) a boy or a girl?	1. YES 2. NO	1. YES 2. NO
09. In what year was (he/she) born?	YEAR	YEAR
010. Is (he/she) living with you?	1. YES 2. NO	1. YES 2. NO
011. In what religion (was (name) raised/is (name) being raised)? Is it Protestant, Catholic, Jewish, some other religion or none? (IF PROTESTANT) What specific denomination is that? (PROB)	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:
012. INTERVIEWER CHECKPOINT SEE 09	1. CHILD WAS BORN BEFORE 1973-AGE 01 2. ALL OTHERS--NO TO NEXT CHILD	1. CHILD WAS BORN BEFORE 1973-AGE 01 2. ALL OTHERS--NO TO NEXT CHILD
013. What is (he/she)'s religious preference? Is it Protestant, Catholic, Jewish, some other religion or none? (IF PROTESTANT) What specific denomination is that? (PROB)	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:

GO TO NEXT CHILD OR TURN TO P. 74, G11

	c. NAME 03	d. NAME 04	e. NAME 05
	1. MALE 2. FEMALE	1. MALE 2. FEMALE	1. MALE 2. FEMALE
	1. YES 2. NO	1. YES 2. NO	1. YES 2. NO
	YEAR	YEAR	YEAR
	1. YES 2. NO	1. YES 2. NO	1. YES 2. NO
	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:
	1. CHILD WAS BORN BEFORE 1973-AGE 01 2. ALL OTHERS--NO TO NEXT CHILD	1. CHILD WAS BORN BEFORE 1973-AGE 01 2. ALL OTHERS--NO TO NEXT CHILD	1. CHILD WAS BORN BEFORE 1973-AGE 01 2. ALL OTHERS--NO TO NEXT CHILD
	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:	PROTESTANT DENOM: CATHOLIC JEWISH NONE OTHER, SPECIFY:

GO TO NEXT CHILD OR TURN TO P. 74, G11

GO TO NEXT CHILD OR TURN TO P. 74, G11

Figure 2. The RID questionnaire design for questions G11 and G13

E19. Many people have some good friends they feel close to. Who are your good friends (other than your spouse)? Just tell me their first names. (LIST NAMES IN GRID BELOW) (Is there anyone else?) (ENCLOSURE IS USED)

[REPORT E50-E53 FOR EACH FRIEND.]

- Now let's go back and talk about (NAME).
- E50. (NAME) is (male/female)? Is that correct?
- E51. About how old is (NAME)?
- E52. About how many years of school has (NAME) completed?
- E53. What is (NAME)'s religious preference? Is it Protestant, Catholic, Jewish, some other religion, or none?

(IF PROTESTANT): What specific denomination is that?

E49. FIRST NAME	E50. SEX	E51. AGE	E52. EDUCATION	E53. RELIGIOUS PREFERENCE		
				PROTESTANT DENOMINATION	CATHOLIC	JEWISH (SPECIFY)
1.						
2.						
3.						
4.						
5.						

**A**

Figure 3. A, The DAS questionnaire design for question E53. B, The RID questionnaire design for question E53.

E19. Many people have some good friends they feel close to. Who are your good friends (other than your spouse)? So that I can tell them apart in the next few questions, would you please just tell me their first names? (Is there anyone else?) (ENCLOSURE IS USED)

	a. NAME #	b. NAME # 2
E50. (NAME #1, #2, etc.) is male/female? Is that correct?	1. MALE 2. FEMALE	1. MALE 2. FEMALE
E51. And how old is (NAME #1, #2, etc.)?	AGE	AGE
E52. About how many years of school has (NAME #1, #2, etc.) completed?	YEARS	YEARS
E53. What is (NAME #1, #2, etc.)'s religious preference? Is it Protestant, Catholic, Jewish, some other religion or none? (IF PROTESTANT): What specific denomination is that? (NONE)	PROTESTANT DENOM: _____ CATHOLIC _____ JEWISH _____ OTHER, SPECIFY: _____	PROTESTANT DENOM: _____ CATHOLIC _____ JEWISH _____ OTHER, SPECIFY: _____

DO NOT WRITE BEHIND OR OVER THESE LINES. SEE INSTRUCTIONS ON PAGE 50 P. 64, SECTION 7

**B**

The DAS design was judged misleading in view of the probing and writing tasks associated with “Protestant” answers. Although interviewers were expected to write something inside the “Protestant Denomination” answer box, the heading for this column in the DAS grid omitted the “(Specify)” caption present in the “Other” column, which also required something to be written in. At the same time, interviewers may have been encouraged to merely *X* the “Protestant Denomination” box, without further probing, by analogy to the Catholic and Jewish situations. All three religious preferences were elicited from respondents in the same manner—from the initial part of question E53 (or H27)—but only one (Protestant) required that interviewers refrain from *X*ing the answer box in order to probe. For these reasons, we predicted that “Protestant” answers to questions E53 and H27 in DAS ran the risk of going unprobed with just an *X* appearing in the box. These error conditions were identified through questionnaire lookups.<sup>3</sup>

In the RID study (see fig. 3*B*), the problem was solved by providing boxes labeled “Other, Specify” and “Protestant Denom.” Each of these included a line where the answers were to be written in. Thus, the visual cues were made to correspond with the task expected of interviewers.

It will be easy for the reader to verify that “Protestant” responses to questions G11 and G12 in the DAS children’s grid (fig. 1) were also threatened by the same format problems, and that the RID version of these children questions (fig. 2) corrected the threat of unprobed answers as described above.

Compared to the RID study, the DAS questionnaire format generated a higher proportion of Protestant mentions unprobed as to denomination. The percentage of unprobed Protestant mentions for E53 was 8.82 in DAS but only 0.80 in RID. Decreases of comparable magnitude in the proportion of unprobed Protestant responses between DAS and RID were also obtained for questions G11, G12 (G13 in RID), and H27.

### The Relevance of Interviewer Experience

Is quality of design a factor of greater importance in accounting for the performance of inexperienced as opposed to experienced interviewers? To research this question, we generated two effect scores for

3. Lookups were necessary because unprobed Protestant answers were not segregated to a unique coding category but appeared commingled under “Protestant: Type Not Specified” with other answers that were dutifully probed but failed to produce an identifiable denomination name.

each interviewer in each study, one score for each effect of interest: ITEM NA (for questions skipped in error) and PROT-NA DENOM (for unprobed Protestant answers). The score is a count of the number of interviews manifesting the effect of interest from among the first three effect-eligible interviews that each study interviewer conducted. Three interviews is the lowest common denominator in order to standardize on size of interviewer load for the questions of interest. The choice of the earliest interviews from each interviewer's eligible load controls for potential effects of supervisory feedback on later interviews.

Three conditions hold if questionnaire design accounts for differences in effect rates between DAS and RID:

1. The average effect scores for RID interviewers should reflect fewer questions skipped in error and fewer unprobed responses than the comparable scores for DAS interviewers.
2. A comparison of RID versus DAS effect scores for the group of SRC repeater interviewers who worked on both studies should confirm the overall effect trends across the two studies.
3. Within each study, no difference in mean effect scores should be observed between the lesser and the more experienced groups of interviewers. Questionnaire design is expected to affect equally all interviewers who worked on a particular study.

#### QUESTIONS SKIPPED IN ERROR

The distributions of ITEM NA effect scores confirm our expectations about the unimportance of interviewer experience.<sup>4</sup> Proportionately fewer students than SRC interviewers (71 percent vs. 82 percent) completed the first three interviews without making any mistakes, but SRC interviewers who made mistakes repeated the error with greater regularity. As a result, these two DAS groups do not appear to differ in the average number of interviews per interviewer with ITEM NA errors. In RID, all newly trained and 90 percent of experienced interviewers had the first three interviews free of ITEM NA errors. The average number of interviews per interviewer in error does not seem to differ for these two RID groups. In contrast, and underscoring the importance of questionnaire design, a lower proportion of DAS than RID interviewers turned in error-free interviews for these children questions (77 percent vs. 92 percent), and the average number of interviews per interviewer with ITEM NA errors is significantly higher in DAS than RID. The results for SRC repeater interviewers exhibit a

4. Detailed tables for these and all other results presented in the paper are available by writing directly to me.

similar trend across studies—in DAS, 79 percent of these interviewers completed error-free interviews compared to 93 percent in RID.

#### UNPROBED ANSWERS

The results parallel the findings discussed immediately above. There is no association in RID or in DAS between the level of interviewer experience and the average number of interviews per interviewer with at least one unprobed Protestant mention. In contrast, the proportion of interviewers with error-free interviews in RID exceeds those in DAS by over 30 percent (96 percent vs. 65 percent), yielding a significant difference in the mean number of interviews per interviewer with unprobed Protestant mentions across studies. The importance of questionnaire design as an explanatory variable is also confirmed by the comparison involving SRC repeater interviewers—only 64 percent of these interviewers probed all Protestant mentions in DAS, while 100 percent complied with this requirement in RID.

### Discussion

The findings suggest that layout and other graphic cues built into the format of survey questionnaires play an important role in communicating question objectives to interviewers, and that differences in design yield detectable effects. These nonverbal cues can either help or hurt the quality of collected data, as we have shown.

The analysis also shows that, contrary to popular wisdom, interviewer experience does not make up for format deficits in survey instruments. It appears that the situations analyzed in this paper are not associated with interviewing familiarity but with generic human perceptual and cognitive qualities relating to the processing of graphic information.<sup>5</sup>

Careful inspection of completed pretest interviews is always recommended to identify potential design problems, but pretest results do not tell us how to fix the problems, and problems may go undetected if the number of cases or the number of interviewers is small. We need to learn more about the mapping of individual interviewing tasks into specific questionnaire designs in order to identify criteria that can help us choose the optimal version from among alternative designs of the same survey questions.

5. Spence and Lewandowsky (1990) provide evidence of the importance of human graphical perception with regard to interpreting and remembering data from statistical graphics.

There are other reasons besides quality to strive for optimal questionnaire designs. Error reductions translate into cost savings in other survey areas such as data retrieval (for information accidentally missed by interviewers) and training (less time will be required to train and supervise interviewers if the survey instruments are implemented optimally). The cost benefits that optimal questionnaire design can generate in these two areas are not insignificant.

## References

- Babbie, Earl R. 1979. *The Practice of Social Research*. 2d ed. Belmont, CA: Wadsworth.
- Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.
- . 1983. "Mail and Other Self-administered Questionnaires." In *Handbook of Survey Research*, ed. Peter Rossi, James D. Wright, and Andy B. Anderson, pp. 359–77. New York: Academic Press.
- Moser, C. A., and G. Kalton. 1974. *Survey Methods in Social Investigation*. New York: Basic.
- Sheatsley, Paul B. 1983. "Questionnaire Construction and Item Writing." In *Handbook of Survey Research*, ed. Peter Rossi, James D. Wright, and Andy B. Anderson, pp. 195–230. New York: Academic Press.
- Spence, I., and S. Lewandowsky. 1990. "Graphical Perception." In *Modern Methods of Data Analysis*, ed. John Fox and J. Scott Long, pp. 13–57. New York: Sage.
- Sudman, Seymour, and Norman Bradburn. 1983. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.