



Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training

Jacques Billiet; Geert Loosveldt

The Public Opinion Quarterly, Vol. 52, No. 2 (Summer, 1988), 190-211.

Stable URL:

<http://links.jstor.org/sici?sici=0033-362X%28198822%2952%3A2%3C190%3AIOTQOR%3E2.0.CO%3B2-R>

The Public Opinion Quarterly is currently published by Oxford University Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/oup.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

IMPROVEMENT OF THE QUALITY OF RESPONSES TO FACTUAL SURVEY QUESTIONS BY INTERVIEWER TRAINING

JACQUES BILLIET AND
GEERT LOOSVELDT

Abstract In this study we analyze the effects of interviewer training on the quality of responses. Data from a field experiment reveal a number of significant differences between trained and untrained interviewers in terms of nonresponse and the amount of information obtained. For some questions, an interaction effect occurred between interviewer training, the use of tape recorders, and the kind of responses. The effects of training appear to be dependent on the structure of the questions. Since such effects occur primarily with questions that assume a great deal of interviewer activity, it is suggested that they are indeed due to the application of the techniques acquired by training, namely *giving instructions, probing, and feedback*. A brief analysis of the interviewer–respondent interaction supports this interpretation.

It is generally accepted that the task performance of the interviewer can have an effect on the responses to factual survey questions. This awareness of the interviewer's potential for manipulating responses has generated numerous attempts to control interviewer behavior. Task rules (Brenner, 1981a, 1981b) and special interviewing techniques (Cannell, Miller, and Oksenberg, 1981) have been designed in the expectation that their correct application will yield an improvement in response quality. Moreover, efforts have been made to develop instruments for analyzing question–answer sequences in order to have a

JACQUES BILLIET is Professor of Methodology and GEERT LOOSVELDT is a Senior Researcher both at the Department of Sociology of the Catholic University of Leuven, Belgium. This research was funded by the National Foundation of Scientific Research, project no. 2.0050.82. The authors wish to thank Professor Bryan Wilson, Professor Karel Dobbelaere, Dr. Will Dijkstra, Anne O'Brien, Luc Bovens, and the referees for their helpful criticism and suggestions.

better understanding of the actual course of interaction during the interview (Cannell, Lawson, and Hausser, 1975; Dijkstra, Van der Veen, and Van der Zouwen, 1980; Brenner, 1982).

Whether or not the learning of task rules and special techniques by the interviewer have an effect on interview responses can be determined by means of *effect evaluation*. This requires a field experiment and multivariate analysis. However, this kind of analysis provides no information regarding the extent to which trained interviewers actually carry out the learned behavior and no information regarding the specific aspects of task performance to which the response effects can be attributed. Such insight can only emerge from *process evaluation*, involving an analysis of the interviewer–respondent interaction by means of tape recordings of the interviews. Any attempt to improve interviewer behavior must be based on information from both effect and process evaluations.

In this study, we will analyze the effects of specific interviewer training on the quality of the responses to factual survey questions. The questions we deal with are the following: Do trained interviewers obtain fewer nonresponses and more information to questions for which relatively high nonresponse rates and underreporting can be expected? Do trained interviewers actually apply the rules and special techniques they are taught, and does their task performance differ from that of untrained interviewers? In order to answer these questions, both effect and process evaluations were performed (Loosveldt and Billiet, 1984; Loosveldt, 1985a).

Method

DATA AND RESEARCH DESIGN

A field experiment was designed in order to measure the effects of interviewer training on the responses obtained at the interview. The experimental group of respondents were interviewed by specially trained interviewers (see below). The control group were interviewed by interviewers who had only received elementary instructions, as is common in Belgium (see below). About half of the interviews were tape-recorded for process evaluation. Since it is possible that the presence of a tape recorder influences interviewer behavior and therefore also generates response effects, the research design was expanded with the factor *tape recorder* vs. *no tape recorder*. This factor was crossed with the experimental factor *training* vs. *no training* (see Table 1). In this way, it was possible to examine the effect of tape recording on the

interview responses and the interaction effect between tape recording and interview training.¹

The set of respondents ($N = 501$) consisted of a stratified random sample, proportional by municipality, from the population of married women aged between 22 and 55 years, residing in a rural area east and northeast of Leuven, Belgium. This limitation to a rural area avoids the problem of regional variance. The respondents were randomly assigned to the four conditions and to the interviewers. This relatively homogeneous population was selected in order to be able to observe the effect of the experimental factors as purely as possible. The subject of the study, the life situation of married women, determined the choice of the respondent group. This topic was chosen because it facilitates the emergence of a variety of quality indicators.

The questionnaire covered the following issues: family composition, employment of the partners, division of family tasks, evaluation of the residential area, health complaints and medical consumption, social life, political preference, recreation, the family budget, and the sexual relationship with the husband. It also included a social desirability scale. The questionnaire contained both factual and opinion questions.

Given the sometimes contradictory results regarding the influence of "role independent characteristics" of interviewers (Sudman and Bradburn, 1974:117; Delamater, 1974:57), the set of interviewers was also kept as homogeneous as possible: all interviewers were female and had a minimum of interview experience. Of the 62 interviewers who applied, a group of 24 were selected with this need for homogeneity in mind.

TRAINING

From the outset, special care was taken to make sure that the four groups of randomly assigned interviewers did not have any contact

1. The analysis of variance in this study involves a nonorthogonal design, which raises a problem in the partition of the total sum of squares (SS) because the treatment effects and their interaction are correlated. It is thus not possible to determine what fraction of the total SS is caused by what effect (Kerlinger and Pedhazur, 1973:187; Edwards, 1979:162). To resolve this problem, the "experimental design approach" was used (Kerlinger and Pedhazur, 1973:189–193). An additional problem arises with the analysis of variance in our experimental design. According to Dijkstra (1983b) the individual interviewers should be included in the experimental design as a potential source of variation. If the interviewer variance is significant within the experimental conditions, this will always lead to an overestimation of the effect of the experimental variable. In this study, we always checked first for interviewer variance. This involves a design in which the factor "interviewer" (conceived as random) is nested under the experimental training factor. In the event of significant interviewer variance, the pooled mean of squares of the interviewers within the experimental conditions was used to test the effect of training. Of course, if there is no significant interviewer variance, an analysis without taking interviewer variance into account would give an adequate result (Dijkstra, 1983b:182). Significant interviewer variances will be reported when they occur.

with each other, for the true object of the study—research into interviewer behavior—had to be concealed. The training was organized separately for two groups (those with tape recorders and those without).

During the training sessions (five three-hour sessions, see Appendix A) the interviewers watched videotapes and mock interviews and also participated in role-playing sessions. In these sessions three major points were stressed: (1) the *elementary task rules* (Brenner, 1982); (2) accurate and systematic *instructions*, negative and positive *feedback*, and *probing* (Marquis and Cannell, 1969; Cannell, Oksenberg, and Converse, 1979:109–116, 133–135) to promote the role of “faithful respondent” (Weber and Cook, 1972:274–276; Dijkstra and Van der Zouwen, 1978:63; Dijkstra, 1983a:34–35); (3) effective interviewer *reaction behavior* to inadequate respondent behavior (Dijkstra, Van der Veen, and Van der Zouwen, 1980:27; Brenner, 1982:152–163). Since the emphasis in our training was primarily on promoting the role of “faithful” respondent (i.e., a respondent who tries to provide for accurate information), a task-oriented (or impersonal) interview style was chosen. The aim of the rules and techniques was to guarantee sufficient interviewer activity so as to avoid any disadvantages of the impersonal style (Cannell, Oksenberg, and Converse, 1979:42–44). Cannell believes that answering questions is a difficult task, but during the interview the respondent can be taught to perform this task with some degree of accuracy. With this aim in mind we designed our training program.

For each group of untrained interviewers (with tape recorders/without), a three-hour meeting was held. During the meeting, we made a number of administrative arrangements, explained how the questionnaire should be used, and introduced the elementary task rules. As is common in surveys in Belgium, the untrained interviewers were not taught any special techniques, no special attention was paid to effective reaction behavior to inadequate responses, and the elementary task rules were not practised.

The questionnaires used by both the untrained and trained interviewers were the same and had identical printed instructions. All interviewers were instructed to read the printed “instructions to respondents” and all were told to heed the printed “instructions to interviewers” (see Appendix B). During the training sessions, special attention was paid to the use of the two sets of instructions. The trained interviewers were told explicitly to read the “instructions to respondents” word for word, and the “instructions to interviewers” were explained in detail and practiced. In order to familiarize the interviewers with the questionnaire and to enable us to correlate the interviewer and respondent characteristics afterward, the questionnaire was filled out completely by all the interviewers.

Table 1. Planned (in Parentheses) and Actual Number of Interviews in the Four Experimental Conditions

	Training	
	Yes	No
With tape recorder		
Respondents	123 (200)	162 (175)
Interviewers	6 (8)	7 (7)
Without tape recorder		
Respondents	118 (125)	98 (100)
Interviewers	5 (5)	4 (4)

FIELDWORK

The interviews were conducted in several phases so that in principle no interviewers from different groups were at work during the same period, in order to prevent the interviewers from communicating with each other. The interviewing began in mid-October 1981, and at the end of December 80% of the interviews were completed. The remaining 20% were completed by February 1982. Two trained interviewers in the tape recorder group dropped out prematurely and could not be replaced. Therefore, the expected number of interviews for that group was reduced to 150 from the planned 200. Furthermore, two of the trained interviewers (with tape recorder) completed less than 20 interviews because of circumstances beyond their control. Among the untrained interviewers, three replacements were made from our reserve pool (before the start or at the early stages of interviewing). In total, 676 respondents were approached, of whom 25.9% refused to participate.² The total number of interviews for analysis was 501. The refusal rate was 23.3% for the trained and 28.2% for the untrained group.

2. Since there is compulsory voting in Belgium, we could use the electoral registers (per municipality) of the general elections (November 1981), which are exhaustive and up-to-date lists of all adults. The registers include information on age, sex, marital status, and address of each resident of voting age. A stratified random sample of 800 married women between the ages of 22 and 55 (without replacement), proportional by municipality, was drawn. According to the original sampling plan (Table 1), 600 women were randomly assigned to the four conditions and within each condition to the interviewers. Since two trained interviewers dropped out before the start, we could use 250 units for random replacements. In total, 86 women could not be contacted (3 recalls). Of the 676 women approached, 175 refused to participate. Only 7 women did not match the population characteristics. The other 31 units were not used for reasons such as the dropping out of the interviewers before they had conducted 25 interviews.

The comparability of the respondents in the four experimental groups was essential for this study. Since the actual samples (Table 1) could be somewhat biased by the dropout of interviewers and respondents, we were compelled to check the four experimental conditions on a number of respondent characteristics that are presumed to be independent of response effects. Analysis of variance indicated that there was no difference between the experimental groups in age and number of children of the respondents. Log-linear analysis also showed no significant association between the experimental factors and the employment (whether or not) and education variables.³

THE DEPENDENT VARIABLES: QUALITY INDICATORS

The first objective of the study was to determine the extent of the effect of training on the quality of responses to factual questions. The dependent variables are, therefore, indicators of "quality." Assuming that the true value is unknown, an indirect method must be used for evaluating this quality. A number of quality indicators were used,⁴ but in this paper we will confine ourselves to the quality indicator "response completeness." This indicator has two variants. In the first place, the *response rates* may be higher to questions for which a relatively high rate of nonresponse could be expected. Second, the response may contain a greater *amount of usable information* for questions for which underreporting is expected.

From a number of validity studies, it appears that "underreporting rates are related to the elapsed time between events to be reported, to the salience of events for respondents, and to the respondent's perceptions of the social desirability of the events" (Cannell, Miller, and Oksenberg, 1981:396–397). Underreporting increases as the time between an event and the interview increases. Events which are less important to the individual (nonsalient) are reported less completely than those of more importance. If an event is perceived as embarrassing, threatening, or divergent from one's self-image, it is likely to be either not reported at all or distorted in a desirable direction (Cannell, Oksenberg, and Converse, 1979:8–9).

3. For an excellent introduction to log-linear modeling see G.J.G. Upton, 1978:46–118.

4. Other quality indicators for this study are (1) the willingness to give socially undesirable responses, (2) construct validity with regard to the expected relationships between political preference, religious participation, and membership in Catholic organizations, (3) the kind of information (specificity), and (4) the absence of congruence of the opinions of the interviewer with those of the respondent. In this article we will restrict ourselves to the two quality criteria mentioned—response rates and amount of information—since they are exclusively related to the factual questions and to the so-called perceptual questions, which we consider factual questions insofar as they query for information about objective facts (Kalton and Schuman, 1982:42). In a later report, we will discuss the other quality indicators, which are also related to opinion questions.

Therefore, a list of items can be drawn up for which it may be assumed that an increased amount of information (more reporting) is more valid. The amount of information is operationalized in two ways: the reporting of an event and the number of items reported.

In the effect evaluation, we will analyze the response rates to nine factual questions with relatively high nonresponse rates. Furthermore, we will examine the response completeness for fifteen other factual questions dealing with nonsalient or embarrassing events or with events distant in time. The questions used are listed in Appendix B.

Effect Evaluation: Results and Discussion

RESPONSE RATES

The questions with relatively high nonresponse rates concern the political preference of the respondent (Q 175) and her husband (Q 176), the net monthly family income (Q 239), the amount saved monthly if the respondents actually do save (Q 242), discussion of the sexual relationship with the partner (Q 246), evaluation of the sexual relationship (Q 247), personal sexual life (Q 248), frequency of sexual intercourse (Q 249), and the last sexual contact (Q 250). It was expected that the proportion of nonresponses to these nine questions individually and to all of them together would be significantly lower for the trained interviewers than for the untrained interviewers.

The trained interviewers had fewer nonresponses on each of the nine questions for which high nonresponse rates were expected than the untrained interviewers, but the differences between them are significant ($p < .01$) for only six questions out of the nine (Table 2). It is somewhat surprising that no significant effect was found on the two questions concerning the political preference of the respondent and her monthly family income, in contrast with the questions about the husband's political preference and monthly savings. In total, for all nine questions, the trained interviewers had incomplete information from 33.2% of the respondents, the untrained interviewers from 54.6% ($\chi^2 = 23.25$, $df = 1$, $p < .001$).

Does tape recording in conjunction with the training also play a role? It is possible that the interviewers who know they are being recorded will try harder to obtain more information. The effect of training and of tape recording on nonresponse (the nine questions together) and the interaction between training and tape recording was tested with log-linear models.

In the upper part of Table 3 one can see that there is no marginal effect on the amount of information due to the presence of the tape

Table 2. Percentages of Respondents with No Responses to Nine Questions for Which a High Nonresponse Rate Was Expected, by Training Condition

Question	Training		χ^2	df	p
	Yes (N=241)	No (N=260)			
Political party of respondent	6.64	8.08	0.37	1	n.s.
Political party of husband	16.60	22.69	2.93	1	<.10
Income	5.81	8.46	1.31	1	n.s.
Savings	10.37	21.92	12.18	1	<.001
Discussion of sexual relations	0.41	3.85	6.85	1	<.01
Evaluation of sexual relations	2.07	8.46	10.00	1	<.01
Evaluation of sexual intercourse	4.56	16.54	18.64	1	<.001
Last sexual intercourse	5.39	15.00	12.40	1	<.001
Frequency of sexual intercourse	10.37	14.62	2.04	1	n.s.

Table 3. Percentage of Respondents with Incomplete Information (I) to the Nine Questions Together by Experimental Condition and the Model Selection of the Log-linear Analysis

Tape Recorder (R)	Training (T)		Total
	Yes	No	
Yes	35.8	49.4	43.8
No	30.5	63.3	45.4
Total	33.2	54.6	

Model Selection	df	L.R. Chi-square	Prob.
TRI (saturated model)	0	0.00	1.0
TR, TI, RI	1	4.55	.0329
Difference by omitting TRI	1	4.55	.0329

NOTE: Base N of each cell is given in Table 1.

recorder. But there is evidence for a conditional effect: the interaction effect between training (T), tape recorder (R), and completeness of information (I) is apparent from the log-linear analysis.⁵ The interaction term TRI cannot be omitted, so only the saturated model is adequate. This means that the relation between response completeness and tape recording varies depending on the training of interviewers.⁶

Presumably, the presence of the tape recorder stimulated the untrained interviewers to convince a number of unwilling respondents to answer all the questions. The trained interviewers were as successful, but for them the presence of a tape recorder did not have any effect.

THE AMOUNT OF INFORMATION OBTAINED

For questions concerning nonsalient or socially undesirable events or events distant in time, we may assume that a larger amount of information is an indication of better response quality.

The questions about advantages and disadvantages of the residential area (Q 68, Q 69), about memberships in organizations (Q 174), and about receiving magazines (Q 181) were considered nonsalient. Strictly speaking, the two questions mentioned first are perceptual questions, but they can be considered factual insofar as they envisage information about objective facts (Kalton and Schuman, 1982:42). Buying on credit (Q 241) and drunkenness of the husband (Q 287) are socially undesirable.

The items about health complaints and medical consumption, which were largely inspired by the study of Cannell, Oksenberg, and Converse (1979), concern events distant in time. They were covered by an open question about health complaints during the preceding three weeks (Q 73), a closed question about consultation of medical experts (Q 73.2), an open question about the use of medications for each cited complaint (Q 73.3), a closed question about 14 health complaints (Q 74 or Q 75), and an open question about medical consumption during the preceding three weeks (Q 76–79). The health complaints may certainly refer to nonsalient or embarrassing events.

As expected, more respondents reported complaints to the trained interviewers than to the untrained ones (Table 4). The difference is largest for the open question which was asked first and for which the

5. Since the TR relationship in the experimental design was fixed, it must be included in all models (this effect was not tested). This also applies for all the following log-linear analyses (see Upton, 1978:81).

6. The effect parameter $\tau_{111}^{RIT} = 0.904$ (standardized: -2.125 , $p < .05$). This means that the conditional associations between tape recording and completeness of information differ according to training condition. $\tau_{111}^{RIT} = 0.942$ and $\tau_{111}^{RIT} = 1.153$ (R = tape recorder present; T = training; I = no training; I = complete information).

Table 4. Percentage of Respondents Who Reported an Item, by Training Condition

Item	Training		χ^2	df	p
	Yes (N = 241)	No (N = 260)			
Complaints (open)	56.9	34.5	24.24	1	<.01
Complaints (closed)	88.8	78.8	8.33	1	<.01
General practitioners	31.7	26.2	1.59	1	n.s.
Specialists	11.6	12.0	0.00	1	n.s.
Dentists	5.0	7.3	0.80	1	n.s.
Medications	60.3	46.2	9.37	1	<.01
Advantages of res. area	96.7	93.1	2.60	1	n.s.
Disadvantages of res. area	63.9	66.9	0.38	1	n.s.
Membership	45.2	45.8	0.01	1	n.s.
Magazines	88.0	85.4	0.51	1	n.s.
Credit buying	34.2	21.7	8.94	1	<.01
Drunkenness	17.8	15.4	0.52	1	n.s.

vague concept of “complaints” was very broadly defined in the preceding instructions for the interviewer.⁷ This effect can therefore be obtained by correctly applying the instructions. The same is true for the question about medication (Q 79: medication or not), which is also open to a broad or a strict interpretation according to how the instructions are applied. However, the effect on the closed complaints question has nothing to do with the definition of complaints, since the response alternatives offered defined the complaints in the same way for each respondent (Schuman and Presser, 1981:85). This question requires adequate interviewer activity such as reading each of the 14 complaints and noting the responses. Furthermore, the application of the written instructions (“tell the respondent once again that she should calmly think about it . . .”) could increase the reporting of complaints that otherwise would not have been recalled. Contrary to expectations, there is no effect on the reporting of consultation of medical experts during the preceding three weeks (Q 76–78).

7. On the open health complaints question, about 10% of the respondents in both the trained and the untrained condition reported a complaint that could be considered embarrassing: infections of the womb or bladder, hot flushes, breast surgery, breast amputation, hemorrhoids, menopause, infection of the mammary glands, and difficult childbirth (Cannell, Oksenberg, and Converse, 1979:104, 177). Therefore, contrary to our expectations, there was no effect on the reporting of this kind of information.

There is also no significant difference in the reporting of advantages and disadvantages of the residential area, of memberships in organizations, and of the receiving of magazines. Contrary to expectations, the number of respondents reporting disadvantages is slightly higher in the untrained group. With regard to the advantages and disadvantages of the residential area, we shall see that the reporting in the untrained group can be affected by the use of leading probes (see below). This may also apply to the other items. The question on credit purchases (socially undesirable behavior) reveals an effect in the expected direction, but the reporting of drunkenness of the husband shows no effect.

In order to examine the effects of the training and of the tape recorder, log-linear models were tested. For each question, we checked whether the inclusion of the association between training and response is necessary and sufficient to describe the data adequately. An additional interaction effect between training, tape recording, and response was only found for the question regarding the use of medication.

The second operationalization of the amount of information deals with the number of items reported.⁸ The analysis can be based either on all respondents or on only those who reported at least one item. We prefer the latter because four questions contain a kind of filter (Q 73, Q 181, Q 241, and Q 274). The interviewers had to probe for more information in the second part of the question if the respondent gave an affirmative answer and passed the filter. For these and other questions (Q 68, Q 69) the trained interviewers were taught to give positive feedback (see Appendix A) and to probe if an event was reported. The expected training effects on getting more information after a filter or an initial answer can thus be stated more precisely if we look only at those respondents who passed the filter and reported at least one event.

Because of the role learning (the role of "faithful" respondent) that presumably took place during the early portion of the interview and because of the greater amount of interviewer activity (probing and feedback) during the particular questions, greater reporting of items was expected in the trained interviewer group. Since higher proportions of respondents reported an event in the trained condition (Table 4), we could perhaps more easily find significant results if the analysis were based on all the respondents.

From Table 5 (based on the respondents who reported at least one event), we learn that there is a training effect on the number of com-

8. The number of respondents who reported more than one episode of drunkenness (Q 287) or who reported several consultations of a specialist or a dentist is too small for an analysis of variance. The analysis was carried out for the three questions about medical experts together (Q 76–78). The duration of medications (Q 79) can hardly be considered an indicator of the amount of medications; therefore we do not use that question here.

Table 5. Mean Number of Reported Items in the Four Conditions with the Significance Levels of the Main Effects and Their Interactions (Analysis of Variance)

Tape Recorder (R): Training (T):	Yes		No		<i>p</i>		
	Yes	No	Yes	No	T	R	T × R
Complaints (open)	1.79 (69)	1.31 (60)	1.56 (69)	1.13 (30)	<.01	<.05	n.s.
Complaints (closed)	4.18 (106)	4.04 (129)	4.80 (108)	3.63 (76)	<.05	n.s.	<.10
Consultations of medical experts	1.62 (51)	2.00 (67)	2.19 (47)	1.85 (54)	n.s.	n.s.	n.s.
Advantages of residential area	2.80 (121)	2.22 (151)	2.36 (112)	1.91 (91)	<.01	<.01	n.s.
Disadvantages of residential area	2.03 (83)	1.47 (116)	1.57 (71)	1.43 (58)	<.01	<.01	<.05
Memberships	1.32 (61)	1.48 (78)	1.37 (48)	1.35 (42)	n.s.	n.s.	n.s.
Magazines	2.98 (105)	2.36 (138)	2.85 (107)	2.38 (84)	<.01	n.s.	n.s.
Credit purchases	1.50 (40)	1.21 (32)	1.52 (42)	1.29 (24)	n.s.	n.s.	n.s.

NOTES: The number of respondents is given in parentheses. For the test procedure, see footnotes 1 and 9.

plaints reported.⁹ With the open complaints question, the tape recorder also played a role, but the tape-recorder effect is lost if the analysis is based on all the respondents. The interaction of the amount of information with training and tape recorder in the closed complaints question can be explained by the fact that the effect of the training emerges best when the tape recorder is absent. Nevertheless, it is curious that the combined effects of training and tape recording are not cumulative. Contrary to our expectations, there is no effect on the registration of the number of consultations of medical experts.

The effect on reporting advantages and disadvantages of the residential area and the effect on the reporting of magazines are again in line with our expectations: training and, presumably, the awareness of be-

9. Analysis of variance with the factor interviewer nested under the experimental factor training shows only significant interviewer variance for the number of advantages ($F = 2.88$, $df = 21$, $p < .01$) and disadvantages ($F = 1.80$, $df = 21$, $p < .05$) of the residential area. However, the training effects on the number of advantages ($F = 9.49$, $df = 1$, $p < .01$) and disadvantages ($F = 8.24$, $df = 1$, $p < .01$) remain significant. None of the other questions involves significant interviewer variance. Therefore, in those cases, the model is tested without taking interviewer variance into account (see note 1 above).

ing checked (tape recording) probably stimulate a greater amount of interviewer activity (probing and feedback). However, the tape recorder can have a beneficial effect on the interviewers' behavior, provided they know what is expected of them. This should be the case particularly in the trained condition. It is somewhat surprising that we do not find such an interaction pattern in the closed complaints question. The effect of training possibly increases when the question structure leaves more room for autonomous interviewer activity. Therefore, the absence of an effect with the question concerning memberships can be accounted for since probing is included in the question structure in such a way that training could hardly have any effect. The training effect on the reported amount of credit purchases is significant ($p < .05$) when the analysis is based on all the respondents.

Evaluation of the Question-Answering Process

Did the trained interviewers actually apply the rules and special techniques taught? Are the response effects due to the trained interviewers' performance? These are relevant questions because it is possible that, analogous to the concept of "good" subjects (Weber and Cook, 1972), there are "good" interviewers who, following training, try to obtain responses in accordance with the researcher's expectations but without applying the techniques and rules (e.g., by leading probing). Response completeness or more information can only be accepted as a quality indicator if it is brought about by *adequate* interviewer behavior. The only way to evaluate this is by an analysis of the interviewer–respondent interaction. Such an interaction analysis was carried out by means of transcribed tape recordings of 270 usable interviews (115 by 6 trained and 155 by 7 untrained interviewers).

The recorded data permit an evaluation of the question–answer sequences in the two experimental conditions (Brenner, 1982; Dijkstra, Van der Veen, and Van der Zouwen, 1980; Loosveldt, 1985b) and it was also possible to analyze some interaction characteristics for those questions that show response effects. In this paper we will confine ourselves to those questions, although the analysis can also be done for the other questions. For five questions that are used in the effect evaluation¹⁰ the following interaction characteristics concerning the special

10. The five questions deal with the advantages (Q 68) and disadvantages (Q 69) of the residential area, health complaints (Q 73), reception of magazines (Q 181), and credit purchases (Q 241). Those five questions showing response effects are those for which the task rules and special techniques (giving instructions and feedback, nondirective prob-

techniques learned and the appropriate task rules were coded: reading the question as it is worded, applying the printed instructions, giving positive feedback if the respondent answered adequately and provided more usable information (see Appendix A), nondirective probing, and avoiding leading probes.

READING THE QUESTION

The most elementary interview rule states that the question must be read as it is written in the questionnaire. This was virtually always the case for the trained group. For the untrained group, deviations occurred somewhat more frequently. For more than 90% of the respondents interviewed by untrained interviewers, four out of the five questions were read correctly. Inadequate reading occurred primarily with the question about disadvantages of the residential area. This question was adequately read in the trained condition in 92.4% of the cases compared to 78.4% in the untrained condition ($\chi^2 = 10.5$, $df = 1$, $p < .01$). The most common deviation was that the question was not asked at all, probably because disadvantages had already been mentioned by the respondent in answer to the question about advantages of the residential area, which immediately preceded it, so that the question was apparently considered unnecessary.

GIVING INSTRUCTIONS

We checked the question about health complaints for the application of the introductory instructions to the interviewer which were included in this part of the questionnaire. The printed instructions contained a broad definition of the term "complaint" and explicit instructions for the respondent (think carefully, let your memory work, etc.). There are clear differences between the two groups of interviewers. In the trained group, the instructions and the definition of complaint were given in 89% of the cases. In the untrained group, those instructions were almost completely neglected. The untrained interviewers did not use either the definition of "complaint" or the explicit instructions in over half of the cases (59.3%). For more than a quarter of the respondents (27.1%), the definition of complaint was given but not the explicit instructions. For 10% no definition was given but the instructions were given adequately.

ing) are of particular significance. We do not use the questions for which higher response rates were found, assuming that this effect was brought about not so much by question-related interviewer activity as by role learning ("faithful respondent") during the early part of the interview.

It is obvious from the other questions that neither the trained nor the untrained interviewers gave task instructions spontaneously if they were not explicitly stated in the questionnaire, although the trained interviewers had been instructed to do so.

FEEDBACK

The trained interviewers were instructed to give adequate feedback. This means giving positive feedback if the respondent answered the question correctly and provided more usable information. They were urged to give slightly negative feedback or no feedback at all if the respondent answered too quickly or gave a negative answer (or nonresponse). In the untrained condition, almost no positive feedback was given if the respondent answered correctly or gave more information. For the questions about health complaints, magazines, and buying on credit, the trained interviewers gave positive feedback to about 19% of the respondents. Concerning the question about buying on credit, however, this feedback was *inadequate* in 40% of the cases where it was used (i.e., positive feedback on negative answers). The trained interviewers occasionally gave feedback for those questions, but this feedback was often inadequate.

PROBING

An important aspect of interviewer behavior is probing. Adequate probing—i.e., asking for more information without suggesting an answer and repeating the question after a negative response—varied strongly according to interviewer group. For each of the five questions, significantly more respondents in the trained group were probed than in the untrained one (Table 6). Taking the five questions together, the mean number of probes per respondent was 6.13 for the trained group against 2.56 for the untrained one ($F = 159.2$, $df = 1$, $p < .01$).

However, the mean number of *leading* probes for those questions was .30 for the trained group against 1.08 for the untrained group ($F = 28.3$, $df = 1$, $p < .01$). The most striking differences were found in the questions about advantages and disadvantages of the residential area. In the trained group 9.8% (Q 68) and 7.9% (Q 69) of the respondents were probed inadequately (leading probe); in the untrained group, this happened with 22.3% (Q 68) and 23.9% (Q 69) of the respondents. These differences were due to the fact that untrained interviewers tended to give more sample responses after an initial negative answer. During the training, it was stressed that this should not be done. The leading probes were possibly responsible for the absence of some expected response effects in the lower part of Table 4.

Table 6. Percentage of Respondents Who Were Probed Adequately on Five Questions, by Training Condition

Question	Training		χ^2	df	p
	Yes (N = 115)	No (N = 155)			
Complaints (Q 73)	80.7	52.6	22.3	1	<.001
Advantages (Q 68)	89.2	38.1	70.1	1	<.001
Disadvantages (Q 69)	86.7	41.1	56.4	1	<.001
Magazines (Q 181)	66.3	48.6	8.0	1	<.01
Credit purchases (Q 241)	56.3	21.6	31.7	1	<.001

Conclusions

The results of our study show that interviewer training of five half-day sessions generates a number of visible response effects. For a number of questions for which nonresponses and underreporting were expected, the trained interviewers succeeded in obtaining more complete information. In a number of cases, however, the training seemed to have had no effect. The occurrence of the training effects turned out to be dependent on the question structure, with greater effects for those questions that required more activity from the interviewer. The opportunities to convey to the respondents what is expected of them by means of instructions, to probe, and to give adequate positive and negative feedback occur particularly with open questions and with the presentation of lists of items. The effects of the training must, therefore, be related to question structure.

In a number of cases, we found a significant effect for the use of the tape recorder. This effect is ascribed to the interviewers' awareness that they were being checked, which led them to perform better. This assumes that the interviewers know what is expected of them, which is mainly the case for trained interviewers. However, it is also possible that the effects of training and taping are not cumulative.

The results from the process evaluation in the present study indicate that the trained interviewers generally applied the rules they were taught during the training, and that their performance obviously differed from that of the untrained interviewers. Nevertheless, the trained interviewers also had their shortcomings: they did not provide enough adequate feedback or spontaneous task instructions. The training could certainly be improved in this respect.

Appendix A: Short Description of the Training

The training was conducted in five sessions of approximately three hours each.

First meeting: introduction to and explanation of the content of the research. Only the substantive project "life situation of married women" was presented. The true objective of the project was concealed. Reference was made to the scientific and social relevance of the project in order to stimulate cooperation. The questionnaire was then presented in full, and the various parts discussed. The interviewers were asked to indicate the parts they expected to have difficulty with.

Second meeting: the theory of the interview and interviewing. Presentation of the task rules and observation of interview situations on videotape ("Verantwoord vragen en stijl van de interviewer" by W. Dijkstra and J. Van der Zouwen, Vrije Universiteit Amsterdam). Discussion of the situations. The aim was to give the interviewers insight into the various aspects of the interviewer task and its importance. Task rules: reading the questions as they are worded; asking every question that applies to the respondent; using prompt cards and other stimulus instruments where required; probing nondirectively only; making sure that the answer is correctly understood and that it is adequate; never answering for the respondent; not giving directive information; not giving unrelated information; repeating actions if asked for by the respondent; giving nondirective clarifications in accordance with the question objectives; reacting nondirectively after an inadequate answer (Brenner, 1982:139–140).

Third meeting: practice of the task rules and learning of adequate reactions to respondent behavior by means of role playing and self-made videotapes with various interview situations. Acting nondirectively after an inadequate answer: probe, repeat question, repeat instruction, or clarify.

Fourth meeting: learning and practice of the special techniques: giving instructions, probing, positive and negative feedback (Cannell, Oksenberg, and Converse, 1979). Learning to apply the printed "instructions to interviewers" and to read the printed "instructions to respondents" literally. Furthermore, learning to encourage the respondents (think calmly, be accurate, etc.) on other occasions (spontaneously giving instructions). Avoiding leading probes. Giving positive feedback (approval: "um-hmm," "all right," "O.K., thanks," etc.) if questions are answered adequately (e.g., more information) and negative feedback (disapproval: "you answered that quickly," etc.) if they are answered inadequately (e.g., no response, too quickly) (Cannell, Oksenberg, and Converse, 1979:109–116). Guidelines for introduction at the door. The interviewers carried out a test interview.

Fifth meeting: discussion and evaluation of the test interviews. Concrete arrangements with regard to refusals and replacements. Distribution of the questionnaires and, finally, the filling in of the entire questionnaire by each interviewer.

Appendix B: Question Wordings (Original in Dutch)

[no preceding printed instructions]

Q 68: Can you give me a number of advantages of the residential area here?
(Probe)

Q 69: Can you give me a number of disadvantages of the residential area here?
(Probe)

Instructions for the interviewer (not to be read to the respondent)

A number of questions now follow about health complaints that the respondent may have had recently. By complaints we understand both physical complaints and complaints deriving from anxiety, fear, or tensions that occur in their environment and because of which the respondent felt less well. Injuries or lesions due to accidents are also to be considered complaints. It is very important that the respondent answers these questions as completely as possible. Therefore, tell the respondent once again that she should calmly think about it, should allow her memory to function, and should answer according to her situation.

Q 73: Have you had any complaints about your health in the last three weeks?

No. Go to Q 75

Yes. What?

(for each complaint ask the following three questions and note the responses in the table)

Q 73.1: Can you describe your complaint?

Q 73.2: Whom have you consulted for this complaint? This could be (A) nobody, (B) general practitioner, (C) doctor-specialist, (D) dentist, (E) physical therapist, (F) psychotherapist, other (who?)

Q 73.3: Have you taken medication for this complaint? (Yes/No)

Q 74: You have mentioned a complaint (some complaints) that you have experienced in the last three weeks. We know from experience that many people often forget a complaint. So I will read a number of complaints to you. Will you try to remember if you have had any of these complaints in the last three weeks? (Go to the list of Q 75)

Q 75: You just said that you have had no complaints in the last three weeks. However, we know from experience that many people often forget a complaint as soon as it is over. So I will read a number of complaints to you. Will you try to remember if you have had any of them in the last three weeks?

(List with 14 complaints:)

Less appetite than normal, shortness of breath, bad or sweet taste in the mouth, nervousness, heart palpitations or thumping in the heart area, muscle pain, feeling of fatigue, sore throat, stomach complaints, headache, sleepiness and drowsiness, back pain, dizziness, listlessness (yes/no).

Instructions for the respondent (to be read to the respondent)

(If no complaints:)

It could be, even though you had no complaints, that you still consulted a general practitioner, a specialist, or a dentist for one reason or another.

(If complaints but did not consult a doctor:)

It could be that you have consulted a doctor for other reasons than for the complaints you just mentioned.

(If complaints and consultation:)

You have already mentioned a consultation (a number of consultations) of a doctor for your complaints. I will now ask about the precise number. It could also be that you have gone to a doctor for other reasons than for the complaints you have just mentioned. Will you count these consultations?

- Q 76: When you think back over the last three weeks, how often have you consulted a general practitioner for yourself?
- Q 77: Again thinking back over the last three weeks, how often have you consulted a specialist for yourself?
- Q 78: How often have you been to the dentist for yourself in the last three weeks?
- Q 79: Think back one last time over the last three weeks, how many days have you taken medications?

[no preceding printed instructions]

- Q 174: Are you a member of any of the following organizations?
- Q 174.1: Women's organization (No/Yes: which?)
- Q 174.2: Athletic organization (No/Yes: which?)
- Q 174.3: Political organization (No/Yes: which?)
- Q 174.4: Religious organization (No/Yes: which?)
- Q 175: Which political party do you prefer?
- Q 176: Which political party does your husband prefer?
- Q 181: Do you receive any magazines in your home? (No, go to Q 182) (Yes, which? Do you read them yourself? No/Yes)

Instructions for the interviewer (not to be read to the respondent)

In order to have a full insight into the situation of married women, we would like to ask the respondent a number of questions on financial matters concerning the family and about her relationship with her husband. Some of these questions the respondent will probably not answer readily. You must assure the respondent that the answers are strictly confidential and that information on these subjects is also very important for the study. Stress that the questions must be answered in all sincerity and that they concern the situation of the respondent.

- Q 239: When you look at the following scale of incomes (show Card 18), in what category does the total net income of your family per month, everything included, fall? Answer by giving the letter of the appropriate category.
[A: less than 20,000 fr., B: 20,000 to 29,999 fr., etc.]
- Q 241: Do you now have things in your home that have been purchased on credit since your marriage? (No/Yes)
Yes: Which? Are you still making payments on them? (repeat these questions per item)
- Q 242: Do you save regularly? (No/Yes)
Yes: How much do you generally save per month?
Under what form do you save? In the form of bonds, a savings or deposit book, a term account, by inscription to public loans or other forms (several answers are possible).
- Q 246: How easily can you talk with your husband about sexual matters in your marriage? (read the answer possibilities A through F and show Card 19)
A: they are never discussed;
B: with difficulty, I cannot find the right words;
C: with difficulty, generally my husband does not discuss them;
D: with some difficulty;
E: rather easily discussed;
F: easily.
- Q 247: Would you say about experiences of sexual intercourse you have had up to the present (read response possibilities A through D and show Card 20)
A: that they have been very unsatisfying and for you unpleasant experiences;
B: that they were not unpleasant but as a whole not satisfying;
C: that, taken as a whole, they were pleasant experiences;
D: that, taken as a whole, they were particularly beautiful experiences for you.
- Q 248: Do you have the feeling that your personal sex life up to the present (read response possibilities A, B, and C and show Card 21)
A: has generally been worse than that of other people;
B: has generally been just as good as that of other people;
C: has generally been better than that of other people.
- Q 249: Would you like to have much more, more, about the same, less, or much less sex than you do at present?
- Q 250: When was the last time you had sexual intercourse? (read response possibilities A through F and show Card 22). Please answer by indicating the appropriate letter.
A: less than three days ago;
B: less than one week, but longer than three days ago;
C: less than two weeks, but longer than one week ago;
D: less than three weeks, but longer than two weeks ago;
E: less than one month, but longer than three weeks ago;
F: longer than one month ago.

[no preceding printed instructions]

Q 287: It sometimes happens that people drink too much alcohol and become drunk. How many times did this happen to your husband during the last three months?

References

- Brenner, M. (1981a)
 "Patterns of social structure in the research interview." In M. Brenner (ed.), *Social Method and Social Life*. London: Academic Press.
- (1981b)
 "Skills in the research interview." In M. Argyle (ed.), *Skills and Work*. London: Methuen.
- (1982)
 "Response effects of 'role-restricted' characteristics of the interviewer." In W. Dijkstra and J. Van der Zouwen (eds.), *Response Behaviour in the Survey Interview*. London: Academic Press.
- Cannell, C. F., S. A. Lawson, and D. L. Hausser (1975)
A Technique for Evaluating Interviewer Performance. Ann Arbor: Survey Research Center, University of Michigan.
- Cannell, C. F., P. V. Miller, and L. Oksenberg (1981)
 "Research on interviewing techniques." In S. Leinhardt (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.
- Cannell, C. F., L. Oksenberg, and J. M. Converse (1979)
Experiments in Interviewing Techniques: Field Experiments in Health Reporting 1971–1977. Ann Arbor: Survey Research Center.
- Delamater, J. D. (1974)
 "Methodological issues in the study of premarital sexuality." *Sociological Methods and Research* 3:30–61.
- Dijkstra, W. (1983a)
Beïnvloeding van Antwoorden in Survey-Interviews. Utrecht: Elinkwijk.
- (1983b)
 "How interviewer variance can bias the results of research on interviewer effects." *Quality and Quantity* 17:179–187.
- Dijkstra, W., L. Van der Veen, and J. Van der Zouwen (1980)
 "Effects of interviewer style on response behaviour in survey-interviews." Paper presented at the symposium "Uses of the Research Interview," University of Surrey, 17 December.
- Dijkstra, W., and J. Van der Zouwen (1978)
 "Role playing in the interview: Towards a theory of artifacts in the survey-interview." In R. F. Geyer and J. Van der Zouwen (eds.), *Sociocybernetics*. Leiden: Nijhoff.
- Edwards, A. L. (1979)
Multiple Regression and the Analysis of Variance and Covariance. San Francisco: W. H. Freeman and Co.
- Kalton, G., and H. Schuman (1982)
 "The effect of the question on survey response answers: A review." *Journal of the Royal Statistical Society* 145:42–57.
- Kerlinger, F. N., and E. J. Pedhazur (1973)
Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston.
- Loosveldt, G. (1985a)
Het Effect van een Interviewertraining op de Kwaliteit van Gegevens bekomen via het Survey-Interview. Doctoral thesis. Leuven: Faculteit der Sociale Wetenschappen.

——— (1985b)

“Evaluatie-onderzoek naar het effect van een interviewtraining.” *Tijdschrift voor Sociologie* 6:355–381.

Loosveldt, G., and J. Billiet (1984)

Experimenteel Onderzoek naar de Voorstructureerbaarheid van een Interviewsituatie. Leuven: Sociologisch Onderzoeksinstituut.

Marquis, K. H., and C. F. Cannell (1969)

A Study of Interviewer–Respondent Interaction in the Urban Employment Survey. Ann Arbor: Survey Research Center, University of Michigan.

Schuman, H., and S. Presser (1981)

Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Sudman, S., and N. M. Bradburn (1974)

Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.

Upton, G. J. G. (1978)

The Analysis of Cross-Tabulated Data. New York: John Wiley and Sons.

Weber, S. J., and Th. D. Cook (1972)

“Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid influence.” *Psychological Bulletin* 77:273–295.