

Interviewer Judgments about the Quality of Telephone Interviews

John Tarnai¹ and M. Chris Paxson²
Social & Economic Sciences Research Center¹
College of Business²
Washington State University

Abstract

Telephone interviewers are sometimes asked to judge the quality of interviews that they conduct, and to describe problems that they may have encountered during the interview. This kind of data is rarely reported, but is a core component of the kind of survey quality that is the theme of this conference. In this paper, we present the results of asking interviewers to judge the quality of telephone interviews for 12 different telephone surveys conducted during the last several years. Interviewers rated each interview (excellent, good, fair, poor, inadequate), assessed whether respondents were interested in the survey topic, whether they had strong views on the topic, and the reasons that the quality of information obtained was less than excellent (language, hearing problems, interruptions or distractions, poor phone connection, competency, infirm, intoxication, respondent was rushed, respondent did not take interview seriously, R did not understand meaning of some questions, R was offended by interview, someone else was listening). Additionally, we present the results of an analysis that compares the data from higher quality versus lower quality interviews, with a goal of identifying the consequences of including less than excellent interviews in a survey dataset. We also compare these interviewer judgments of quality by type of respondent and survey topic. Finally, we discuss the implications these results may have for interviewer training, data analysis, and design of telephone surveys.

Keywords: Telephone Survey, Interviewer Judgment, Survey Quality, Telephone Interviewer Ratings

1. Introduction

Who knows more about the quality of a telephone interview than the interviewer him or herself? In telephone surveys we generally are pleased to get as many completed interviews as we can without really considering the quality of those interviews. But interviewers usually have a pretty good idea of how well or badly an interview went, and consequently can

gauge the quality of the survey data that were collected.

A variety of things can occur during a telephone interview to affect the quality of the data collected, both from the perspective of the respondent and from the perspective of the interviewer. Respondents may feel rushed to complete the interview because it may be interrupting something else that the respondent wants to do. Respondents may feel irritated about being asked to do the interview and not inclined to take the interview seriously. Some respondents may be infirm, intoxicated, or have health problems that make it difficult for them to concentrate on the questions and provide proper answers. There may be problems with the telephone equipment so that the respondent has difficulty hearing the questions that are being asked. Respondents may have difficulty understanding some of the questions asked in the survey. Respondents may choose not to answer questions that they do not understand or that they find objectionable. Sometimes respondents may be reluctant to answer sensitive questions if they feel that others in the household may be listening to their responses. For these and a variety of other reasons, the survey data collected by telephone interviewers may vary in its quality.

We could not find any research that had been conducted about interviewers' own assessments of the quality of the survey data they collected. However, there is a substantial literature on interviewer effects (Biemer and Lyberg, 2003; Groves, et al. 2004) on the quality of survey data.

2. Research Questions

In the telephone surveys we have conducted over the past few years, we have made it standard practice to ask interviewers to rate the quality of the interview they have just conducted, and if the quality was less than excellent, to indicate which of several reasons explains the problems. Despite collecting these data for numerous telephone surveys over the years, we have never seriously examined the data or what

interviewers were telling us about the data they collected.

Our main purpose in this paper is to examine interviewer assessments of the quality of telephone interviews for 12 telephone surveys conducted during the past few years. Table 2.1 below describes the 12 surveys included in this study. The main questions explored in this study include:

- What percent of completed interviews were rated as less than excellent?
- What were the main reasons given by interviewers for rating the quality of interviews less than excellent?
- Is there anything in the survey data that corroborates the lower quality ratings?
- Do any characteristics of lower quality interviews cut across all surveys?

Table 2.1 Number of Completed Interviews for the 12 Surveys Included in this Study

Survey	Description	Completed Interviews
1	Household survey about substance use and abuse	6,660
2	Household survey about access to legal resources	1,887
3	Survey of sexual behaviors and experiences	1,194
4	Household survey about curbside recycling	4,354
5	Household survey about air quality opinions	1,101
6	Survey of problems with financial lending institutions	1,361
7	Household survey of political opinions	1,761
8	Survey of women’s experiences with sexual violence	1,325
9	Household survey of opinions about privatizing utilities in Kansas	2,061
10	Household survey of opinions about privatizing utilities in Wisconsin	1,661
11	Survey of parent’s experiences with child care	1,208
12	Survey of child care provider’s characteristics and opinions	286
Total		24,859

The overall goal of the study was to determine whether asking interviewers to rate the quality of interviews is a useful thing to do, and whether it can tell us anything about the quality of the telephone interviews in our survey datasets.

3. Methodology

The data for this paper come from 12 telephone surveys conducted over the past few years, primarily household surveys on topics described in Table 2.1. The surveys included in this study varied by topic, length of interview, type of respondent, and interviewer characteristics as well. While most of the surveys had both male and female interviewers, one of the two surveys (#8) dealing with sexual experiences had only female interviewers. The interview length varied from less than 10 minutes to over 25 minutes.

At the conclusion of each interview, after the interviewer hung up the phone, they were asked to rate the quality of the interview, and then indicate which of 13 reasons were applicable to the interview, if it had been rated as less than excellent:

3.1 Interviewer Questions

- Q. How would you (the interviewer) rate the quality of the information obtained in this interview?
1. Excellent - no problems at all => END
 2. Good - a few problems but overall quality is good
 3. Fair - a number of problems but overall acceptable
 4. Poor - many problems, overall quality open to question
 5. Inadequate -- interview was terminated by interviewer
- Q. What were the reasons the interview was less than excellent?
- A. Not in R's native language.
 - B. Hearing -- lots of background noise.
 - C. Interruptions or distractions.
 - D. Poor phone connection.
 - E. Lack of mental or physical competency to respond.
 - F. Infirm -- too old, weak, ill, etc.
 - G. Intoxication.
 - H. R was rushed.
 - I. R didn't take study seriously.
 - J. R didn't understand meaning of some questions.
 - K. R was offended by the interview.
 - L. Someone was listening in, so R may not have been truthful.
 - M. Some other reason.

(IWR Choices are; 1=yes a reason, 2=not a reason)

These questions were worded exactly the same way for each of the 12 surveys. However, one survey on

substance use and abuse issues (survey #1), included one additional question, described below, which provided another indication of the adequacy of the interview.

3.2 Additional Interviewer Question

{interviewer: the researchers would like to conduct follow-up interviews with willing respondents who might have more info on this topic. in your opinion, would you say this respondent would make a good candidate for follow-up interviews? would you say this respondent . . .

1. Seemed to have a story to tell
 2. Seemed interested in the topic
 3. Had strong views on the topic
 4. Would NOT make a very good candidate for follow-up
- D. Don't know

While interviewers were trained about how to answer this question, some did not take it seriously. We found a number of instances where interviewers simply skipped answering the question. However, we assume that the majority of interviewers were being honest in their assessments of the quality of the telephone interviews.

4. Results

Figure 4.1 below displays the average percent of interviewers giving ratings of excellent, good, fair, poor, and inadequate to the 12 telephone surveys included in this study.

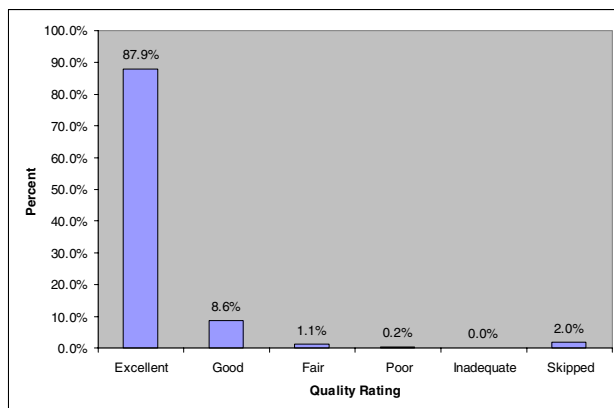


Figure 4.1: Average Rating of the Quality of Telephone Interviews for 12 Telephone Surveys

An average of 2% of interviewers skipped answering the question, and the majority of interviews (almost 88%) are rated as excellent. Only 10% of interviews on average were rated as less than excellent, and most of these (8.7%) were rated as good. Very few were rated as fair (1.1%) and only a handful were rated as poor (0.2%).

As can be seen in Tables 4.1 and 4.2 below, there is some variation in the interviewer ratings of the quality of the interviews, but not a great deal. Except for surveys 11 and 12, all the rest had between 83% and 96% of interviews rated as excellent. Only a very small percent of interviews in each survey were rated as poor (less than one percent) and only slightly more were rated as fair (less than 2%). On the whole, interviewers gave fairly high ratings to the interviews they completed, and identified fewer than 5% of interviews as being less than excellent or good.

Table 4.1 Interviewer’s Quality Ratings for Six Telephone Surveys

Survey #→	1	2	3	4	5	6
	DRG2	TEJF	FOX2	CURB	AIR1	DFIV
Excellent	83.7%	95.4%	85.8%	93.7%	90.8%	90.5%
Good	11.1%	3.7%	4.5%	4.5%	7.6%	8.3%
Fair	1.6%	0.6%	0.6%	0.6%	1.0%	0.7%
Poor	0.2%	0.1%	0.1%	0.2%	0.5%	0.5%
Inadequate	0.1%	0.2%	0.0%	0.0%	0.0%	0.0%
Skipped	3.4%	0.0%	9.0%	1.0%	0.0%	0.0%
Total N	6660	1887	1194	4354	1101	1361

Table 4.2 Interviewer’s Quality Ratings for Six Telephone Surveys

Survey #-->	7	8	9	10	11	12
	CS04	OCVA	PRKS	PRWI	BRAP	BRAC
Excellent	92.6%	96.8%	84.6%	83.3%	76.2%	65.7%
Good	6.5%	2.8%	14.0%	15.1%	14.5%	10.8%
Fair	0.9%	0.2%	1.3%	1.4%	2.6%	2.8%
Poor	0.1%	0.1%	0.2%	0.2%	0.7%	1.4%
Inadequate	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Skipped	0.0%	0.2%	0.0%	0.0%	5.9%	19.2%
Total N	1761	1325	2061	1661	1208	286

When interviewers rated an interview as less than excellent, they were then asked to indicate which of ten reasons explained why they gave a rating of less than excellent. Interviewers were instructed to indicate

as many of these reasons as they felt applied to each interview.

Figure 4.2 displays the average percent of interviewers indicating each type of problem as a reason that interviews were rated as less than excellent for the 12 surveys included in this study. The top six problems most frequently indicated were “Respondent did not understand meaning of some questions” (20.5%), “other reasons” (20.4%), “hearing problems” (11.2%), “interruptions or distractions” (9.3%), “interview not in respondent’s native language” (9.2%), and “infirm” (6.8%). The remaining seven problems were all indicated less than 5% of the time on average over the 12 surveys.

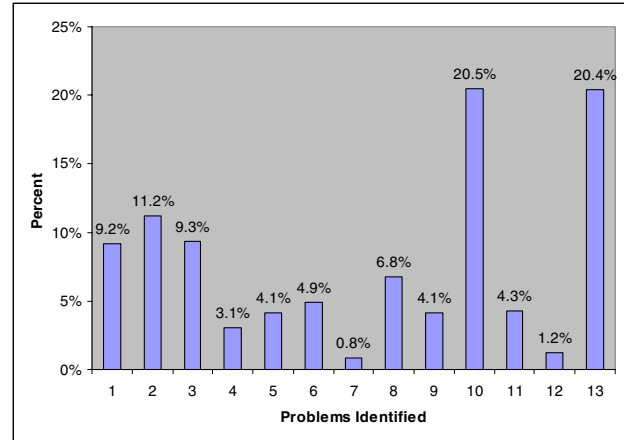


Figure 4.2: Percent of Problems Identified for Telephone Interviews for 12 Telephone Surveys

Table 4.3 Percent of Problems Identified in the First Set of Six Telephone Surveys

Survey #	1	2	3	4	5	6
1. Interview not in R's native language	24.6%	3.8%	9.3%	4.6%	6.2%	1.4%
2. Hearing	8.5%	10.9%	7.6%	12.0%	9.3%	13.4%
3. Interruptions or distractions	6.5%	7.7%	7.6%	5.0%	7.5%	14.1%
4. Poor phone connection	3.0%	1.6%	1.7%	2.0%	2.2%	2.9%
5. Lack of mental or physical competency to respond	5.5%	5.5%	2.5%	7.0%	3.1%	4.0%
6. Infirm	3.9%	4.9%	0.0%	8.6%	5.8%	6.2%
7. Intoxication	1.1%	1.1%	1.7%	1.2%	0.4%	0.7%
8. R was rushed	6.2%	7.1%	2.5%	6.6%	8.8%	6.9%
9. R did not take interview seriously	4.6%	6.0%	4.2%	5.6%	4.4%	3.6%
10. R did not understand meaning of some Qs	18.0%	16.4%	25.4%	23.9%	25.7%	25.4%
11. R was offended by interview	3.3%	10.9%	6.8%	3.0%	3.5%	6.2%
12. R may have been untruthful-someone else listening	1.8%	0.5%	2.5%	0.6%	0.4%	0.4%
13. Other reasons	13.1%	23.5%	28.0%	19.7%	22.6%	14.9%
Number of Interviews	1314	183	118	498	226	276

Table 4.4 Percent of Problems Identified in the Second Set of Six Telephone Surveys

Survey #	7	8	9	10	11	12
1. Interview not in R's native language	3.2%	5.8%	0.8%	0.9%	6.3%	0.8%
2. Hearing	15.7%	10.5%	13.8%	14.0%	10.1%	7.8%
3. Interruptions or distractions	11.2%	12.8%	11.1%	11.3%	15.5%	8.5%
4. Poor phone connection	4.4%	3.5%	3.6%	3.9%	3.3%	3.1%
5. Lack of mental or physical competency to respond	0.8%	3.5%	3.0%	2.2%	4.7%	0.8%
6. Infirm	6.4%	3.5%	6.0%	5.4%	0.9%	3.9%
7. Intoxication	0.0%	2.3%	0.6%	0.7%	0.2%	0.0%
8. R was rushed	6.8%	5.8%	8.5%	8.8%	4.9%	3.9%
9. R did not take interview seriously	3.6%	2.3%	3.6%	3.9%	2.8%	2.3%
10. R did not understand meaning of some Qs	15.3%	15.1%	18.8%	18.5%	28.4%	19.4%
11. R was offended by interview	4.4%	3.5%	1.9%	2.0%	2.1%	33.3%
12. R may have been untruthful-someone else listening	1.6%	7.0%	1.1%	1.1%	0.2%	0.0%
13. Other reasons	26.5%	24.4%	27.2%	27.3%	20.4%	16.3%
Number of Interviews	249	86	637	557	426	129

As the data in the tables above show, while there is variation in the percent of interviewers selecting each reason for the less than excellent rating, there is also quite a bit of consistency across the 12 surveys. Respondents not understanding the meaning of some questions is a main reason for poorer quality interviews in all the 12 surveys, as is interruptions or distractions, and hearing problems. On the other hand, problems with intoxicated respondents is not given very often as a reason for poor quality interviews in all 12 surveys. Despite the differences in the content of the 12 surveys and in types of respondents, there is much similarity in these percents, suggesting some uniformity in how often these kinds of problems occur in telephone surveys.

For one of the surveys (#2) interviewers were asked whether or not the respondent would make a good candidate for follow-up interviews. Rating an interview as excellent did not guarantee that the respondent would also be a good candidate for follow-up, as 17% of excellent interviews, were also rated as “would NOT make a very good candidate for follow-up.” On the other hand, 67% of interviews rated as poor, were also rated as “would NOT make a very good candidate for follow-up.” Among interviews rated as excellent, good, or fair, almost 50% of respondents were rated as “seemed interested in the topic,” with very little variation in this rating.

We were interested in assessing the extent to which there might be other indicators in the surveys themselves to support these quality ratings. To that end, we conducted additional analyses by looking at the total number of missing items for respondents in five of the surveys, and comparing the average number of missing items for each quality rating. We did not do this for all the surveys, mainly because the remaining surveys included many branching items, which made it difficult to differentiate missing items due to legitimate branching and those due to respondents.

Table 4.4 Average Number of Missing Responses by Quality Rating for Each of Five Surveys

Survey #-->	2	3	5	9	10
Excellent	19.96	22.86	6.64	2.19	1.37
Good	20.71	26.78	8.95	3.82	2.93
Fair	24.43	42.86	11.09	6.23	5.29
Poor	19.00	59.00	15.83	4.25	3.25
Minimum Missed	8	12	1	0	0
Maximum Missed	46	60	40	28	27

As indicated by the results in Table 4.4, there is an inverse relationship between quality ratings and the average number of missing items in interviews. Interviews rated as “excellent” had the lowest average number of missing items, whereas interviews rated as “fair” or “poor” had the highest average number of missing items. Thus, it seems that the quality ratings given by interviewers were at least partly affected by the number of times respondents could not or would not answer questions in the survey.

Another confirmation that the quality ratings were related to the number of missing items is that we found a significant difference between the average number of missing items between respondents who had been rated as “not understanding the meaning of some questions”, and those who had not received this rating. The results of this analysis for surveys number 9 and 10 are presented in Table 4.5.

Table 4.5 Comparisons of the Average Number of Missing Items by Whether or Not Interviewers Indicated that Respondents Did or Did Not Understand the Meaning of Some Questions

Survey #9	Yes	No	F	p
N	120	202		
Mean	5.53	3.13	24.21	<.001
Survey #10				
N	103	179		
Mean	4.76	2.23	20.63	<.001

The average number of items missing for these two surveys is not large, but the differences between those interviews for which the interviewer indicated the respondent did not understand the meaning of some questions, and other interviews, are significant. Respondents who did not understand the meaning of some questions missed almost twice as many questions on average as other respondents, and this was true for both surveys. We ran this same analysis for two other surveys (2 and 5), but did not find a significant difference, because the sample sizes were small. However, the average number of missing items was in the same direction, with respondents who did not understand the meaning of some questions having a greater number of missing items.

For all of the surveys in this study, interviewers had the option of writing in other reasons for giving an interview a less than excellent rating. Below are just a few examples of the kinds of “other” reasons given by interviewers for rating an interview as good, fair, poor, or inadequate:

- I think she was paying very close attention to both the survey and the TV at the same time.
- The lady had a lot of things going on at one time; she had two kids she was trying to give baths and she was cussing because too many things were going on at one time.
- The R kept rushing me to finish and threatening that he would quit the interview if I did not hurry up.
- She kept cutting me off before I could finish the question.
- She really didn't give me too many problems it's just that she wanted to answer the questions in her own way half the time and not use the choices that I gave her.
- Respondent would not answer questions using the answer choices provided in the survey. She talked about everything other than the survey, including her kids, work, city council, and many, many other topics. She was a nice lady, but she did not take the questions seriously and she interrupted me to talk about what she wanted to talk about.
- The respondent was mad that I was reading from a script on the screen. He wanted the study to be personalized for his individual needs. He asked to speak to a supervisor and when he did, he still was not happy, so the supervisor decided to terminate the interview.

5. Discussion and Conclusions

The job of an interviewer is not an easy one, and most try to do their best to complete interviews with as many respondents as possible. It has been said (We're not sure by whom) that a large part of an interviewer's job is to train the respondent in how to answer survey questions. Interviewers who receive proper training tend to try hard to get respondents to answer all the questions in a survey and stay on task. As Groves (1989) notes, most interviewer training tries to give interviewers guidelines for proper behavior, which can then be applied to the real life situations that interviewers face. But, trainers are not always successful in this effort.

What this means is that not all completed interviews are of equal quality. However, we keep all completed interviews, regardless of the quality, because they are so important to a good response rate. As we saw in this study, most interviews are problem free. Almost 90% of interviews were rated as excellent, and most of the rest were rated as good. Fewer than 2% were rated as fair or poor.

Because there is so little variation in the quality ratings given to interviews, and because there are so few really poor quality interviews, we were unable to detect any relationship to demographic variables. We attempted analyses that looked at the demographic characteristics of interviews that were rated as less than excellent, but could find no significant relationships. Poor interviews were equally common among male and female respondents and among old as well as young respondents.

There are differences, but how can we ascertain what they are and what they mean for survey research? Probably a main implication of poorer quality interviews is that they seem to have a greater number of missing items than interviews rated as excellent. And, the worse the quality, the more missing items there will be in the survey.

Another implication of this research is that interviewer ratings can tell us quite a bit about what is working and what is not working with a survey. The finding that the main reason for a poor interview is that respondents apparently don't understand the meaning of some questions, is quite enlightening. This tells us that we need to write better survey questions to make them understandable to more people. As Conrad and Schober (2000) report, respondents can interpret questions quite differently than survey designers intend. They suggest that a less standardized interviewing style and a more conversational style may improve respondent comprehension of questions, and may lead to better data quality.

Interruptions and distractions are another main reason for poor quality interviews. We need to find ways to help interviewers deal with this when it occurs, perhaps by recommending they make an appointment for a callback at a more convenient time. Dillman (2000) however, suggests that we evaluate our instructions for interviewers carefully, so we don't end up with unintended response effects.

We have not tried to examine interviewer differences in quality ratings. However, this is something that could be considered. As Stokes and Yeh (1988) have described, interviewers may bring their own personal

biases to the data they collect. Good interviewers must also demonstrate quite a bit of patience with difficult respondents, and sometimes the best course of action is to end the interview before all the questions have been asked. When the respondent clearly does not understand the questions, or expresses frustration with the interview or the questions, then rather than continue with the interview, we allow interviewers to call over a supervisor to make a decision about terminating the interview.

Finally, requiring interviewers to answer questions about the quality of the survey is beneficial because it makes them think about the interview process and what produces a good or a bad interview. Moreover, interviewer ratings provide valuable information to survey researchers about problems with a survey or questionnaire. Research by Groves and Magilavy (1986) suggests that traditional indicators of interviewer quality (response rate, productivity, and supervisory evaluations) may not be good indicators of survey quality.

Acknowledgements

We want to thank the telephone interviewers of the Social and Economic Sciences Research Center at Washington State University who worked on the surveys described in this paper over the past several years. We also want to thank the telephone supervisors and the other data collection staff that contributed their time and energy to all of the surveys mentioned in this paper.

References

- Biemer, P. P. & Lyberg, L. E. (2003) Introduction to Survey Quality. New York: Wiley-Interscience.
- Conrad, F. G. & Schober, M. F. (2000) Clarifying question meaning in a household telephone survey, Public Opinion Quarterly, Vol 64: 1-28.
- Dillman, D. A. (2000) Mail and Internet Surveys (2nd Ed.) New York: Wiley.
- Groves, R. M. (1989) Survey Errors and Survey Costs. New York: Wiley.
- Stokes, L. & Yeh, M. Y. (1988) Searching for causes of interviewer effects in telephone surveys, In: Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L. II, Waksberg, J.

(Eds.) Telephone Survey Methodology. New York: John Wiley & Sons.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004) Survey Methodology. New York: Wiley-Interscience.

Groves, R. M. & Magilavy, L. J. (1986) Measuring and explaining interviewer effects in centralized telephone surveys, Public Opinion Quarterly, Vol 50: 251-266.