Monitoring Field Interviewer and Respondent Interactions Using Computer-Assisted Recorded Interviewing: A Case Study

Susan Mitchell[1], Kristine Fahrney[1], Matthew Strobl[1]

[1] RTI International, 3040 Cornwallis Road, Research Triangle Park, North Carolina 27709-2194

Abstract

A major challenge of managing field staff is the lack of direct information about what takes place during the interview. Computer audio-recorded interviewing (CARI) is a technology that allows laptops to record the verbal exchange between interviewer and respondent. It is a useful tool for monitoring adherence to protocols and interview authenticity. This paper describes the use of CARI in the Study of Community Family Life, an in-person survey of 4,000 households in urban areas. We discuss the methods for question and case selection, and interviewer and respondent acceptance of the technology. We also summarize the type and magnitude of interviewer behavior problems we observed. Finally, we compare CARI with telephone verifications and field observations for monitoring data quality. The paper provides an in-depth picture of interviewer behavior that is typically hidden from view.

Key Words: Computer audio-recorded interviewing, CARI, interviewer behavior coding

1. Background

Computer audio-recorded interviewing (CARI) is used to record respondent-interviewer interactions during in-person interviews and, more recently, telephone interviews. In field studies, recording is controlled by the laptop computers that are used to administer the interview, and recording is turned on and off without the knowledge of the interviewer or the respondent. Thus, while both parties are aware that recording will take place, they do not know when it will occur or which questions will be recorded. CARI recordings are transmitted along with survey data to the home office where they are reviewed by quality control specialists. CARI is used to study interviewer adherence to survey protocols, to provide detailed feedback to interviewers about their performance, and to detect and deter interviewer falsification. It is also used to study sources of measurement error, such as data capture error (Mitchell et al., 2008).

CARI was first demonstrated as a viable technique for quality control monitoring in field surveys on the National Survey of Child and Adolescent Well-Being, conducted by RTI and sponsored by the Administration for Children and Families (Biemer et al., 2000). CARI can be implemented in Blaise (Thissen and Rodriquez, 2004) and other questionnaire languages. Prior research has shown that CARI can be cost-effective compared with reinterviewing techniques for quality control (Biemer et al., 2002; Witt and Wang, 2007). While CARI is not a new technology, its uses are still being discovered and exploited in field surveys, and more recently, in telephone surveys (Thissen et al., 2008).

This study is based on data from the Study of Community Family Life (SCFL). The SCFL is part of the Evaluation of the Community Healthy Marriage Initiative that is also sponsored by the Administration for Children and Families and conducted by RTI International. The goal of the evaluation is to measure the community impacts of intervention programs designed to improve marriage and relationship quality, and family and child well-being in low-income communities. Data were collected from about 4,000 respondents between September 2007 and March 2008.

2. CARI Implementation

CARI has been in use at RTI since 1999 and the computing infrastructure that supports it is well established. The project team customized the CARI review system for the SCFL by specifying the questions to record and the method for selecting cases, as described below. As with other means of interview verification, CARI was implemented on a sample basis because of the expense involved in reviewing all completed interviews.

2.1 Question Selection
We selected three closed-ended questions and four open-ended questions for recording. The closed-ended items were recorded for all completed interviews while the open-ended items were recorded only if they were asked (as determined by the instrument's skip logic). Therefore, each completed interview included between three and seven CARI audio files.

The three closed-ended items were selected to allow project staff to monitor questionnaire administration at the beginning, middle, and end of the 40-minute interview. These items were chosen based not only on their location within the interview, but also for their structure. Each had a preset recording time of 45 seconds, making it necessary to select items that could be adequately asked and answered in that timeframe.

In selecting the four open-end items, our primary objective was to identify and record questions that required advanced interviewer skills such as probing for clarity and multiple responses. We selected a mix of single-response questions and multiple-response questions, with response lists (not read to respondents) of varying lengths and complexities. Selecting the open-ended items enabled us to monitor interviewer performance where it was most challenged by the need to accurately probe, code, and record responses.

2.2 Sample Selection
The first case that each interviewer completed was selected with certainty for CARI review so that quality control specialists could confirm that the interviewer met essential performance standards. Subsequently, the sampling rate was set at 10 percent. Upon receiving 10 cases for an interviewer, the CARI system randomly selected 1 of the 10 for review. This selection was repeated with each set of 10 or more cases received by the system. Depending on the number of interviews completed and their job tenure, individual interviewers had between 1 and 25 cases selected for review. The overall 10 percent verification rate was considered adequate and affordable for quality control purposes, and is the rate often used with telephone or in-person verifications.

2.3 CARI Monitoring System

When interviewers electronically transmitted completed interviews from their laptops to RTI, the audio files were transmitted along with the survey data. Quality control specialists reviewed the audio files using a web-based monitoring system that controlled access to the cases and files. A bilingual specialist reviewed cases conducted in Spanish. Each audio file had a dedicated review screen that included

- a link to the file itself, which was stored as a .wav file;
- the interviewer ID and case ID;
- the date and time that the file was stored and the date and time the review was conducted;
- the status code of the case, such as selected for review or review complete;
- a space to note technical issues that interfered with the review, such as microphone failure;
- a space to indicate the quality of the audio file on a scale ranging from very poor to excellent;
- a set of performance codes that captured interviewer error and error-free performance; and
- a space for the reviewer to make file-level notes regarding performance.

By clicking on the link to the .wav file, reviewers heard the audio played back via Windows Media. After coding the individual audio files, reviewers proceeded to a case-level screen that provided space for an overall assessment of the interviewer's performance. The reviewer then assigned a final review code, indicating whether the case had errors present.

In general, review time per case ranged from 5 to 15 minutes, depending on the number of audio files, the length of each file, the presence of errors, and the sound quality. The SCFL experienced a very low problem rate with file and sound quality. Only 13 of the 2,821 audio files reviewed had audio ratings of poor or very poor, and only two files were marked as having technical problems. Some files did contain voices that were low in volume or difficult to understand, and these cases generally required more review time.

The CARI system produced a range of summary and detail reports that provided information about the status and results. Standard reports included the following:

- *Status Summary*. This report showed an overall summary of case status, listing the number and percentages of cases selected for review, cases not selected, new cases, and cases already reviewed.
- *Review Completion Summary*. This report showed the total number of cases and the number and percentage of cases reviewed for each interviewer.
- *Performance Summary*. This report presented the frequency of each performance code and category by interviewer.
- *Performance Detail*. This report showed the performance code and category results for each audio file reviewed.

The CARI system also allowed for the creation of custom reports that presented file-level and case-level notes made by reviewers. These reports were used to provide specific feedback and instruction to interviewers, as described below.

2.4 Feedback to Interviewers

Using the CARI monitoring system, interviewer performance was assessed and corrected with little delay. When an error was identified, reviewers contacted the interviewer's field supervisor and, as needed, the interviewer him- or herself. Field supervisors were contacted by e-mail if the errors in question were minor and by telephone if the errors were more severe. Supervisors provided interviewers with both positive and negative feedback. In many cases, the feedback process was completed within 1 business day of CARI review.

3. Findings

3.1 Interviewer Acceptance of CARI

CARI monitoring was introduced and explained during interviewer training. In the context of discussing quality control procedures, interviewers were told that their CARI files would be reviewed for reading errors, recording errors, non-neutral feedback, and unprofessional behaviors. They were further instructed that interviewers who had a high CARI refusal rate would receive additional training and might be subject to disciplinary action.

Interviewers generally raised no objection to CARI monitoring during training or field administration. No interviewers cited CARI monitoring as a reason for resigning from the project. At the training, experienced interviewers confirmed that as long as the interviewers did not behave as though respondents should be concerned, respondents were generally receptive to the procedure. Some interviewers later reacted to corrective feedback about high refusal rates in a defensive manner; however, retraining and role-playing with their supervisors typically lowered their CARI refusal rates.

By the end of the project, the overall CARI refusal rate was only 7 percent. Two thirds of the interviewers had refusal rates that were lower than 10 percent. Only 14 percent had refusal rates of over 20 percent.

3.2 Respondent Acceptance of CARI

After respondents consented to the interview, they were asked to consent to CARI. The text in the informed consent form read as follows:

> We are using a special quality control system on this project. The system runs on the computer and may record what you and I say to each other during random parts of the interview. Neither you nor I will know when the computer is recording what we say. The recording will be reviewed by people at RTI to monitor my work. The recordings will be used only for those purposes, and will be kept confidential. May we use this quality control system during your interview?

About 7 percent of respondents refused CARI consent. As Table 1 shows, there were no significant differences in the tendency to refuse CARI by age, but women were slightly less likely to refuse than men. There were also differences by race and ethnicity, with non-Hispanic Whites and non-Hispanic Blacks being less likely to refuse than Hispanics and other race/ethnic groups.

Table 1: CARI Refusal Rates by Gender and Age

	Refusal Rate
Gender*	
Male	8.4%
Female	6.7%
Age	
18–35	7.0%
36–49	8.0%
Race*	
Hispanic	13.4%
Non-Hispanic White	6.2%
Non-Hispanic Black	3.9%
Non-Hispanic Other	10.6%

*=p < .05

3.3 Interviewer Adherence to Protocols

In this section, we examine interviewer adherence to data collection protocols as observed during our review of the CARI audio files. Each file was coded for the presence of the following types of errors: minor reading deviation, major reading deviation, inappropriate probing, unprofessional behavior, and questionable authenticity. These errors and their incidence are described below.

3.3.1 Minor Reading Deviation

This error consisted of minor errors in the reading and delivery of the question or response choices. This error included two types:

1. Errors of omission in which some text, such as a particular response choice or a given word in the question text, was not read aloud to the respondent.
2. Errors of commission in which the interviewer read aloud text that should not have been read to the respondent, such as response choices for open-ended items.

This error code was selected when the error was deemed unintentional and had a minimal effect on the respondent's comprehension and response to the item. In general, these errors resulted from carelessness or a hurried pace and were often precipitated by a respondent's attempt to answer a question before the interviewer had a chance to read it completely.

3.3.2 Major Reading Deviation

This error type consisted of more substantial errors in the reading and delivery of the question or response choices. To reach the level of "major" deviation, the error had to be intentional or have a strong potential to bias the response. This code was applied when the reviewer determined that the interviewer purposefully changed or ignored text, or intentionally read text that was not supposed to be read aloud.

This code was also applied to cases where the interviewer made a series of errors that, on the whole, represented a substantial disruption to the question. For example, the interviewer might both misread the question and fail to read a response choice; the two errors, while individually "minor," may be "major" when considered together.

3.3.3 Inappropriate Probing
This error type consisted of the use of non-neutral feedback, rephrasing questions, or leading the respondent to elicit a certain answer. Such action had the potential to bias the way the respondent answered and adversely affect data quality. Some common examples included

- personally interpreting an unclear response rather than asking for clarification,
- probing by reading text that was not supposed to be read, and
- rephrasing the question in the interviewer's own words.

3.3.4 Unprofessional Behavior
This error type consisted of inappropriate action or unprofessional behavior by the interviewer. Although this type of error occurred rarely, some examples include

- leaving materials unattended,
- allowing the respondent to handle the computer or view the screen,
- allowing the respondent to control project materials such as the showcard booklet,
- proceeding with the interview despite a nonprivate setting or allowing someone other than the respondent to contribute to the interview, and
- rushing or being rude to the respondent.

3.3.5 Questionable Authenticity
This error type consisted of interviewer actions that suggested falsification or fabrication of the interview, either in whole or in part. This error, also occurring rarely, was coded when the audio file contained only the interviewer's voice, or the sound of typing but no voices. To rule out technical problems as the cause, these cases were subject to further verification by telephone or in person by the field supervisor.

3.3.6 Error Incidence
As shown in Table 2, about 48 percent of the cases had at least one interviewer error. While this incidence may seem high, it is important to remember that between three and seven audio files were recorded for each case, and only one file had to contain an error for the case overall to be marked as having an error. Conversely, a case marked as having an error could have multiple types of errors present.

Table 2: Overall Incidence of Errors

Status	Number of Cases	Percentage of Total
Cases with no errors	308	51.9
Cases with errors	286	48.1
Total Cases	594	

Looking more specifically at the types of errors and their frequency, it is evident that reading errors were by far the most common. The most frequently occurring error was minor reading deviation, made in 39 percent of the cases reviewed. About 82 percent of all cases with errors contained this error type. The error *major reading deviation* was detected in 25 percent of all cases reviewed. It was detected in more than half (54 percent) of all cases with errors.

Table 3: Incidence of Errors by Type

Type of Error	Number of Cases With Error	Incidence of Cases With Error Among Cases With Errors Only (N=286)	Incidence of Cases With Error Among All Reviewed Cases (N=594)	Number of Interviewers Committing Error
Reading—Minor Deviation	233	81.5%	39.2%	83
Reading—Major Deviation	154	53.8%	25.9%	58
Inappropriate Probing	6	2.1%	1.0%	6
Unprofessional Behavior	34	11.0%	5.7%	24
Questionable Authenticity	1	0.3%	0.2%	1

Other errors were far less common. The unprofessional behavior error was detected in 34 cases while inappropriate probing was detected in only six. There was a single detection of questionable authenticity. Investigation of this case uncovered falsification by the interviewer, whose employment was subsequently terminated.

In addition to occurring more frequently, the reading errors were made by a greater number of interviewers. The interviewers, on average, committed minor reading deviations (83 out of 110 interviewers) and major reading deviations (58 out of 110 interviews) between two and three times each. The less frequently occurring errors were made by fewer interviewers and were generally not repeated.

4. Comparison With Telephone Verifications and Field Observations

In addition to CARI, two other quality control procedures were employed on the SCFL, telephone verifications and field observations. Telephone verifications were conducted in English and Spanish by telephone interviewers from RTI's centralized call center. During the verification calls, interviewers asked respondents questions about the interview experience and the interviewer's behavior that sought to confirm that the interview actually took place. Results of the verification calls were used to identify suspect cases that required further investigation.

For the field observations, a field supervisor or other member of the project team traveled to each site to observe a sample of interviewers as they conducted their work. Observers used a checklist to record whether the interviewer followed all the appropriate protocols. Following the observation period, the observer provided the interviewer with both positive and corrective feedback and prepared a written summary.

Each approach had its advantages and disadvantages. Field observations provided project staff with the most information and the greatest opportunity to offer immediate and direct feedback to the interviewer. However, the disadvantage was that the interviewers were acutely aware of the observer's presence and may have altered their behavior as a consequence.

Telephone verification provided less information and, for two reasons, the reliability and usefulness of the telephone verification data was somewhat questionable. First, we were limited to a short set of questions about aspects of the interaction that most respondents were likely to remember (such as whether they were paid the incentive). Second, even when respondents provided a response that indicated a violation of protocol, it was difficult to determine if the respondent was recalling the information accurately given the passage of time. Furthermore, when a respondent would not talk with us at all, it was unclear whether this indicated no knowledge of the interview or simple disinterest. When using the telephone verification data, project staff looked for a pattern of problems across all the interviewer's cases. Another quality control method, such as review of the CARI files or in-person contact, was used to confirm that there was a legitimate performance issue that needed to be addressed.

Compared with telephone verifications and field observations, CARI provided the most reliable information. Because there was no cue indicating when the recording was turned on, the interactions we heard suggested that once engaged in the interview, both respondents and interviewers forgot that their interactions might be recorded. Thus, CARI monitoring avoided the problem of interviewers changing their behavior during periods of observation. In addition, we were not relying on the memories and perceptions of respondents with CARI monitoring. The disadvantage of CARI is that it provided information on interviewers' adherence to only interviewing protocols, not contacting or screening. Moreover, cost constraints made it practical to record and review only a portion of the interview.

The cost of each quality control method varied greatly, assuming use of existing computing infrastructures and support systems. For field observations, including travel expenses and the labor of the observer (amortized over the number of interviews observed), the cost to observe one interview was roughly $150.[1] For telephone verifications, including telephone interviewer and supervisor time, the cost to verify one case was about $20. At roughly $12–$18 per case, including time for the reviewer to listen to the files and provide corrective feedback when needed, CARI monitoring was the most cost-effective quality control method.

5. Conclusions

CARI is a mature technology that provides reliable information about the interviewer-respondent interaction in field surveys. It is a cost-effective quality monitoring tool that provides insight into field interviewer behavior and sources of survey error. Acceptance of CARI monitoring is high by interviewers and respondents alike. In the SCFL, over 93 percent of the respondents consented to CARI use, despite being informed that the survey asked sensitive questions about domestic violence and infidelity. Men were slightly more likely to refuse CARI than women, and Hispanics were more than twice as likely to refuse CARI as non-Hispanics.

CARI monitoring indicated that only about half of the completed interviews were completed without interviewer error on the questions that were recorded. The predominant error type—minor reading deviation—was also the least serious. This error was observed in 39 percent of the cases but carries with it the least potential to bias

---

[1]All costs shown are unloaded, direct costs.

responses. Major reading deviations, potentially more serious, were observed in about one in four of the cases. Inappropriate probing and unprofessional behavior were very infrequent. Moreover, CARI served as a useful deterrent to data falsification. Field interviewers were informed during training that CARI would be used to monitor their performance and establish interview legitimacy. Only one interview of the 594 cases reviewed was later confirmed as falsified.

Compared with telephone verifications, a more traditional method of quality monitoring in field surveys, CARI offers the following advantages:

- the ability to provide more informed and detailed feedback to interviewers to improve their performance,
- the information collected is not subject to recall bias by the respondent,
- greater coverage of cases and less respondent burden, and
- reduced cost.

Compared with field observations, CARI offers the following advantages:

- actual interviewer behavior in the field is reflected, not "observed" behavior, and
- reduced cost.

The disadvantage of CARI is that interview coverage is not complete—for cost control, it is practical to record and review only a portion of the interview. Moreover, portions of the respondent-interviewer interaction that do not involve the laptop (for example, contacting and screening) cannot be monitored.

Based on these findings, we recommend that survey planners consider implementing CARI as the primary quality control method in field surveys. Because of its limitations, it should not be the only method, but used in conjunction with telephone verifications (to verify CARI findings) and field observations, CARI offers excellent coverage of interviewers and cases. There are investments to be made in building the CARI computing infrastructure, but these investments can be amortized over multiple surveys and paid back through reduced quality monitoring costs. In field studies, CARI offers the ability to identify and correct interviewer performance problems at a reasonable cost.

References

Biemer, P. P., Hergert, D., Morton, J., & Willis, W. G. (2000). The feasibility of monitoring field interview performance using computer audio recorded interviewing (CARI). *Proceedings of the American Statistical Association's Section on Survey Research Methods*, pp. 1068–1073.

Biemer, P. P., Mierzwa, F., & Thissen, M. R. (2002). *Comparison of the steady-state costs for CARI vs. reinterview approaches for interview verification*. RTI Technical Report to the U.S. Census Bureau.

Mitchell, S. M., Strobl, M, Fahrney, K., Nguyen, M, Bibb, B., Thissen, R., & Stephenson, W. (2008). Using computer audio-recorded interviewing to assess interviewer coding error. *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 4414–4421.

Thissen, M. R., & Rodriguez, G. (2004). Recording interview sound bites through Blaise instruments. *Proceedings of the International Blaise Users' Conference,* pp. 411–423.

Thissen, R., Fischer, C., Barber, L., & Sattaluri, S. (2008). Computer audio-recorded interviewing, a tool for monitoring field interviewers, and improving field data collection. *Proceedings of Statistics Canada Symposium 2008.*

Witt, M., & Wang, J. (2007). *Traditional reinterview methodology versus the CARI monitoring methodology when measurement error exists.* RTI Technical Report to the U.S. Census Bureau.