

An Examination of Within-Interviewer Continuity and Change in the use of Standardized Interviewing Following Repeated Corrective Feedback

Kristine Fahrney¹, Danielle Cohen¹, Ann Burke¹
¹RTI International, 3040 Cornwallis Road, RTP, NC 27709

Abstract

Although standardized interviewing practices are widely advocated in order to reduce interviewer-related measurement error, audio recordings of interviewers' work reveal interviewers' sometimes persistent tendency to paraphrase or reword questions. This presentation includes interviewer performance data obtained through CARI (computer recorded audio interviewing) monitoring on two separate face-to-face surveys. The first is a mid-sized study that employs approximately 85 interviewers who administer an approximately 60-minute computer assisted interview. Most interviewers were monitored at least 3 times during the 3 ½ month survey production period. Our analysis of this data focused specifically on quantifying the proportion of interviewers who showed improvement in reading questions verbatim after receiving corrective feedback. The second survey employs a small number of interviewers (~13) who administered a complex 90-minute computer assisted interview. Interviewers received feedback 3 times over the course of 6 months. The small size of this team allowed us to capture rich descriptive data on whether there was improvement in their use of standardized interviewing techniques following corrective feedback. A brief anonymous survey administered to these interviewers sheds light on competing priorities (e.g. need to follow protocols vs. need to be mindful of the respondent's time or interest level) and is suggestive of ways that researchers can frame corrective feedback to be more persuasive to field staff.

Key Words: Standardized interviewing, Data Quality, Computer Assisted Recorded Interview (CARI), Interviewer Error

1. Background

Standardized interviewing has been widely advocated to reduce interviewer-related measurement error. Fowler and Mangione (1990) define 4 principles of standardized interviewing: read questions as written, probe inadequate answers nondirectively, record answers without discretion, and be interpersonally nonjudgemental regarding the substance of answers. Though there is considerable variation in the protocols for implementing these and other

principles for standardized interviewing both within and between research organizations (see Viterna and Maynard 2002), standardized interviewing protocols typically do not allow interviewers to reword the question when it is initially presented. As Maynard and Schaeffer (2002) note, reading the question exactly as worded is probably the most fundamental technique of standardization, and is even in Schober and Conrad's (1997) guide for conversational interviewing.

Although reading the survey questions as scripted is generally considered central to protocols for standardized interviewing, in practice, reading errors are rather common. For example, Mangione, Fowler, and Louis (1992) found that interviewers misread 17% of survey items, whereas Lepkowski et al (1998) found reading errors in 31% of survey items. Mitchell, Fahrney, and Strobl (2009) distinguished between minor and major reading deviations when reporting error rates from brief (~5 minute) recorded portions of randomly selected interviews and found that 39% of interviews contained at least one minor reading deviation and 24% contained at least one major reading deviation.

However less is known regarding the extent to which interviewer behavior varies over the course of data collection (although see Olsen and Peytchev 2007) and in response to post-training corrective feedback. It has becoming increasingly common for interviews to be monitored using computer assisted recorded interviewing (CARI), which allows a monitor to later review and critique recorded portions of interviews. This technique facilitates the examination not only of differences in error rates between interviewers, but also change or continuity within individual interviewer's behavior over time.

The current exploratory analysis investigates the influence of corrective feedback on interviewer behavior in two contemporaneous field studies. The analysis focuses on within-interviewer continuity and change in reading errors compared to two other types of errors that are inconsistent with standardized interviewing: biasing remarks, and inappropriate probing. This analysis addresses two questions:

- 1) Do interviewers adhere more closely to standardized interviewing protocols in response to corrective feedback?*
- 2) Under what conditions is feedback most helpful in encouraging standardized interviewing?*

2. Methods

2.1 Study A

The first study utilized a sample from a mid-sized national field study. Approximately 85 interviewers conducted face-to-face interviews with approximately 4,500 respondents over a 3 ½ month data collection period. The ~ 60 minute computer assisted personal interview (CAPI) instrument collected information related to health center patients' care-

seeking behaviors, reasons for seeking care, health status, use of services, satisfaction with care, unmet health care needs, and perceived quality of care.

The Study A analysis utilizes data collected through CARI, a technology that is used with increasing frequency to monitor interactions in survey research. During the pre-interview informed consent presentation, interviewers obtained consent for the CARI process. If consent was provided, up to eight sections of the interview were recorded (depending on the instrument skip patterns) totaling approximately 6 minutes and 15 seconds. The recordings were controlled by the laptop; neither the interviewers nor the respondents received any visible or audible queues to indicate when the recording was in process. In all, 87% of respondents consented to CARI recording.

CARI recordings were reviewed for at least 10% of each Field Interviewer's (FI) completed interviews by a data quality team comprised of 4 individuals. Each data quality review team member received a detailed standardized training to encourage inter-coder reliability. Initially, one of the first two completed interviews was reviewed for each FI, in addition to one randomly selected interview within the first 10% complete. The subsequent cases reviewed were either selected randomly or chosen for review because they were completed after the most recent performance feedback was provided by that FI's supervisor to track performance over time.

Each CARI file reviewed was coded for the presence or absence of various errors. CARI review outcomes were maintained in reports and in an electronic CARI outcomes log and relayed to each interviewer via their Field Supervisor in the form of verbal feedback. The majority of interviewers (55/82) received feedback 3 times during the 3 ½ month data collection period while the remaining interviewers received feedback either 1 (6/82) or 2 (21/82) times.

For the purpose of this analysis we focus on the presence or absence of the following types of errors for interviewers who received feedback 3 or more times (n=55 interviewers for which 391 cases were reviewed): minor reading deviation, major reading deviation, incorrect use of probes, and feedback not neutral. Table 1 includes a description of the types of errors coded into each category.

Table 1: Study A Interviewer Error Taxonomy

Error	Description
Minor reading deviation	The FI does not read questions verbatim, but the changes are minor, and do not alter the meaning of the question.
Major reading deviation	The FI does not read questions verbatim, and the changes are major, and may alter the meaning of the question.
Incorrect use of probes	The FI does not utilize probing correctly.
Feedback not neutral	The FI does not provide neutral feedback.

2.2 Study B and Web Survey

The second study utilized a sample from a small ongoing list-based longitudinal study of ~ 2,000 couples across five states. The interview consisted of a face-to-face ~ 90 minute CAPI survey covering a broad array of areas such as family background, parent-child and parent-parent relationships, drug and alcohol use, employment and income history, physical and mental health and social support, to name a few. Study B's field team was comprised of 12 interviewers of which nine had prior survey administration experience at RTI.

The Study B analysis examines CARI recordings from interviews that took place between March 2009 and March 2010. As with Study A, during the pre-interview informed consent presentation, interviewers obtained consent for the CARI monitoring process. The recordings were controlled by the laptop; neither the interviewers nor the respondents received any queues to indicate when the recording was in process. A large number of items within the survey were recorded using CARI. A team member listened to at least 10% of each FI's CARI cases. A total of 214 cases, each averaging 20 minutes of recordings are included in the present analysis. Importantly, and in contrast to the above described Study A, Study B comprised a lower volume of data collection over a longer period of time. Thus, interviewers on Study A received feedback between 1 and 3 times during the 3 ½ month period of data collection, whereas interviewers in study B each received feedback approximately every 3 months over a 12 month period. Feedback focused on the 5 types of errors¹ most commonly heard via CARI:

- Did not read some or all of response options out loud
- Left off transition statements or part of question
- Completely reworded question
- Probing issues (not probing or non-neutral)
- Bias- biasing remarks or assuming respondent answers

Table 2 provides a comparison of the methods for Study A and Study B.

Table 2: Comparison of Study A and Study B Methods

	Study A	Study B
Field Period	3 ½ months	Multi-year
Mode	CAPI	CAPI
Questionnaire length	60 minute	90 minute
Interview Complexity	Low	High
Interviewing team	85 interviewers	12 interviewers
CARI QC rate	10%	10%
Size of CARI files	~6 minutes	~20 minutes
Frequency of feedback	3 times over 3 ½ months	4 times over 12 months

In addition to providing each FI with detailed feedback regarding administration errors, the research team provided the FIs an opportunity to take part in an anonymous 16 item web survey (to ensure anonymity, IP addresses were deducted by a third party not associated with the study). The purpose of the web survey was to identify interviewer opinions and values that could potentially influence their willingness to conform to expectations for standardized interviewing.

3. Findings

¹ The following two Study B errors, 'did not read some or all of response options out loud' and 'left off transition statements or part of question' could be included in either major or minor reading deviation Study A error codes. While Study B 'Bias- biasing remarks or assuming respondent answers' could be included in one of the following two Study A error codes, 'feedback not neutral' or 'recording error'.

3.1 Study A

The presence or absence of errors in each CARI file reviewed was recorded for each interviewer at three points in time: 1) prior to interviewers receiving any feedback, 2) after interviewers received the first set of feedback and prior to receiving the second set of feedback and 3) after interviewers received the second set of feedback and prior to receiving the third set of feedback. Figure 1 shows how the need for repeated corrective feedback differs by error type. Looking first at biasing remarks and behavior, we see that of the interviewers who received corrective feedback on this issue, only one warning was necessary to eliminate this behavior. Similarly, inappropriate use of probes was an issue that could be corrected with most interviewers after just one warning; only approximately 10% of interviewers who exhibited this error required a 2nd verbal warning. The two reading errors were more likely to require repeated feedback, particularly major reading deviations.

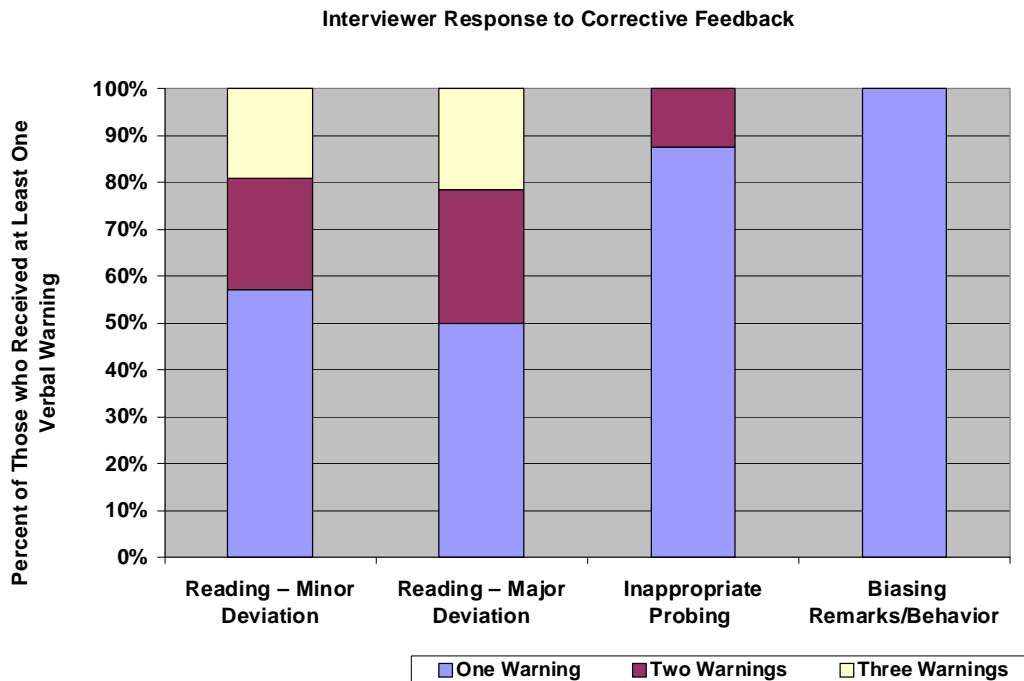


Figure 1: Study A- Interviewer Response to Corrective Feedback

3.2 Study B

As with Study A, the presence or absence of errors in each CARI file reviewed was recorded for each interviewer. However, the extended field period allowed this information to be recorded at four points in time for Study B versus three points in time for Study A. The results for this chart look very similar to the results from study A. Looking at the inappropriate use of probes, we see that of the interviewers who received corrective feedback on this issue, only one warning was necessary to eliminate this behavior. Biasing behaviors was also an issue that could be corrected with most interviewers after just one warning. Nevertheless it should be noted that more than 20% of interviewers who exhibited this error received two verbal warnings. The three reading errors were more likely to require repeated feedback, particularly the error ‘completely rewording the question’.

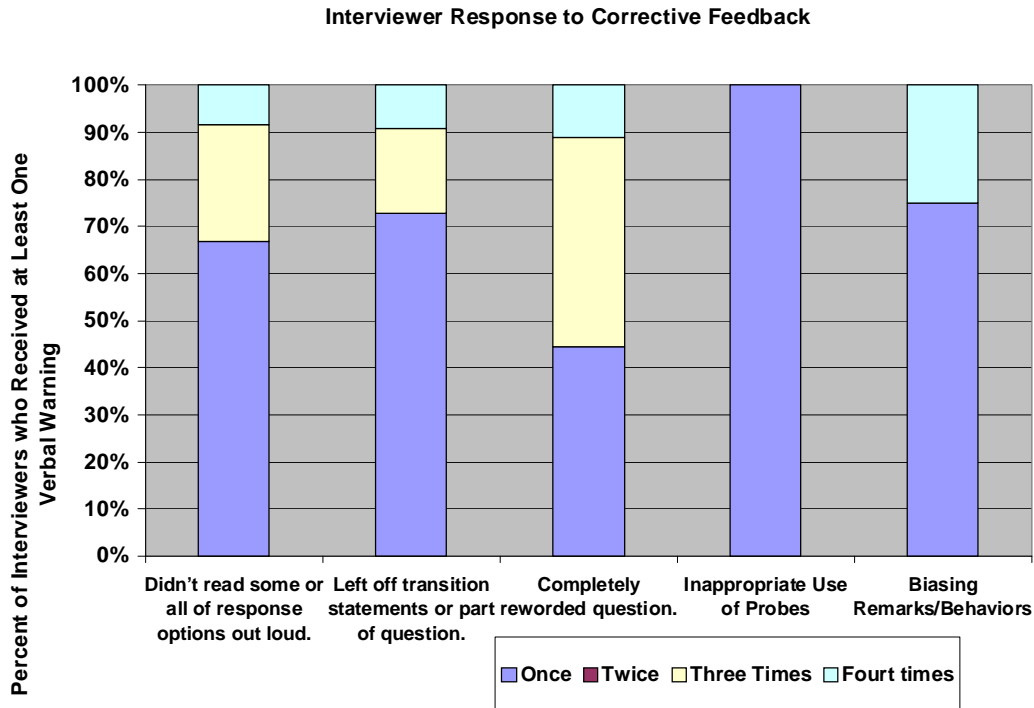


Figure 2: Study B- Interviewer Response to Corrective Feedback

Results of the Study B web survey are presented in tables 3 and 4. The first two closed-ended items presented in Table 3 reflect the FIs’ main opinions about the study, namely that it was too long (55%) and boring (66%) for the respondents. Review of the open-ended follow-up items (for these two items) reveals a trend among FIs that numerous items in the survey are redundant. For example, one FI notes “Many questions seem repetitive to our respondents. Often if the R answers that, for example that the children were removed from the home or live in relative care the survey continues to ask multiple questions about the child.” The reporting of redundancy may explain why interviewers drop parts of the question text and response options more frequently than completely rewording text.

Only one FI (11%) reported that the questions were too difficult for the respondent to understand; in sum, this FI expressed concern about the high reading level of some of the response options. For example, the study uses a standard “Not Applicable” option for some items when a more respondent-friendly version would be “Does not apply”. It should be noted, however, that the “Not Applicable” option is typically presented in all-caps, a convention that requires the FI to determine if that is an appropriate response rather than reading this option aloud to the respondent and requiring him/her to answer.

Table 3- Study B Interviewer Opinions about the Survey Instrument

	Percent Agreement
Survey is too long for respondents	55%
Survey is too boring to respondents	66%
Questions are too difficult for respondents to understand	11%

In line with the main findings presented in Table 3, Table 4 shows that FIs ranked “keeping the interview interesting for respondents” as most personally important whereas “keeping the interview experience interesting for interviewers” least important. Notably, “making sure the interviewers administer the interview exactly as written” was also ranked as most important, possibly reflective of the FIs being aware -- via their training and CARI feedbacks -- that a ‘correct’ answer would be to rank this item as important despite assurances of participant anonymity.

Results from additional (open-ended) items illustrate that the FIs in Study B are clearly aware of study protocols such as the advantages and disadvantages of reading text exactly as written. However, a common theme was FI frustration, on behalf of the respondents, with having to read the same set/series of response options “over and over again”. In one instance, an FI remarked that it was “most annoying” to the respondents. The prevalence of this theme across the FIs’ responses explains – in part – the reluctance to change error 1 (as discussed above) noted by a spike at the third feedback.

Table 4. Study B - Interviewer Ranking of Priorities in Terms of Personal Importance

	Mean	Median	Mode
Keeping the interview experience interesting for interviewers	5.4	5	5
Keeping the interview experience interesting for respondents	2.5	1.5	1
Making the interview easy for the respondent to understand	2.5	2	2
Making sure that interviewers administer the survey in the same way each time	2.7	3	3
Making sure that interviewers administer the interview exactly as written	2.3	2	1

Note: The rank order scale ranged from 1 (most important to you) to 6 (least important to you).

4. Conclusions

Overall, the results indicate that providing corrective feedback can substantially reduce common interviewer errors. However, the effect of corrective feedback on interviewers subsequent behavior varied by error type. On both studies, reading errors were the most common types of errors initially and were also more resistant to change than other types of errors. Results of the Study B web survey suggest that this pattern of findings may reflect interviewers’ competing priorities. While interviewers are mindful that study protocols require standardized interviewing, they are equally concerned with keeping the respondent interested in the survey process. Thus, although interviewers are amendable to making time-neutral changes in their behaviors, such as using neutral rather than biasing probes, they demonstrate reluctance to change their behaviors in ways that would increase the length of the interview, such as reading transition statements, definitions, and lengthy questions, and lengthy response options, as scripted. This suggests that corrective feedback should be framed in a way that addresses interviewers’ priorities (i.e. respondent rapport) rather than survey managers’ priorities (i.e. data quality). Thus, rather than focusing training (and retraining) on the importance of standardized interviewing to data quality, interviewers may be more receptive to reading all text as written if this protocol is presented as a way to allow the respondent the opportunity to more fully understand the questions and as a way to avoid making the respondent’s feel rushed or as if their particular responses are unimportant.

The affect of corrective feedback on interviewer's subsequent behavior also varied between the two studies. While fewer Study A interviewers corrected their behavior after a first verbal warning than the Study B interviewers, considerably fewer interviewers needed a third verbal warning on Study A than on Study B. On Study B, interviewers who continued to make an error after the first corrective feedback all also failed to heed the second corrective feedback, while on study A the proportion of interviewers making a particular error declined substantially with each verbal warning. There are several possible explanations for this pattern of findings. Vieterna and Maynard's (2002) suggest that in order to encourage standardized interviewing, feedback must be both frequent and on an unpredictable schedule. Despite the fact that both studies discussed in this paper involved monitoring at least 10% of each FI's completed interviews, the frequency of this review differed drastically between the two studies. On Study A, feedback was provided on a rolling (frequent and unpredictable basis), and interviewers generally received feedback about their performance within approximately 1 to 2 weeks of completing the interviews that were selected for a quality review. In contrast, on Study B, interviewers received feedback every 2 months about the interviews they had completed since their last feedback. As a result, interviewers on study A may have felt more closely monitored than study B, even though the proportion of work that was monitored was roughly equal on both studies.

Finally, although errors were detected and recorded by a quality control manager and/or team, feedback was provided to the FIs by field supervisors. It is unclear whether the supervisors from both teams placed the same importance on this type of feedback. While the influence of interviewer characteristics on data quality and cost efficiency in survey data collection has been a frequent topic of study, very little is documented about supervisor characteristics that are associated with having the most effective interviewing teams. While neither the Study A nor the Study B sample size was large enough to address this issue, this is a potentially fruitful topic for further examination. Replicating the present analysis with a larger interviewing team would also facilitate more sophisticated data analysis such as calculating survival curves for particular types of errors.

References

- Fowler, F. and T. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park: Sage.
- Lepkowski, J. Couper, M., Hansen, S., Landers, W., McGonagle, K., and J. Schlegel. 1998. CAPI instrument Evaluation: Behavior Coding, Trace Files, and Usability Methods. Proceedings of the Survey Research Methods Section, ASA.
- Mangione, T., Fowler, F., and T. Louis. 1992. Question Characteristics and Interviewer Effects, *Journal of Official Statistics*, 8 (3) 293-307.
- Maynard, T. and N. Schaeffer. 2002. Standardization and Its Discontents. In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, eds T. Mangione, H. Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen. New York: John Wiley & Sons.
- Mitchell, S., Fahrney, K., and M. Strobl. 2009. Monitoring field interviewer and respondent interactions using computer-assisted recorded interviewing: Proceedings of the Survey Research Methods Section, ASA.
- Olson, K. and A. Peytchev. 2007. "Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes." *Public Opinion Quarterly*. 71: 273-286.
- Schober, M. and F. Conrad. 1997. Does Conversational Interviewing Reduce Survey Measurement Error? *Public Opinion Quarterly* 61: 576-602.

Viterna, J. and T. Maynard. 2002. How Uniform Is Standardization? Variation Within and Across Survey Research Centers Regarding Protocols for Interviewing. In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, eds T. Mangione, H. Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen. New York: John Wiley & Sons.