Using Standardized Interviewing Principles to Improve a Telephone Interviewer Monitoring Protocol

Douglas Currivan
Elizabeth Dean
Lisa Thalji
RTI International

Abstract

A critical challenge in monitoring telephone interviewers is accurately recording objective indicators of the degree to which interviews follow standardized behaviors. Without a complete and accurate record of the number and type of specific interviewer behaviors, monitors cannot provide precise feedback to interviewers on standardized interviewing techniques. Objective measurement also facilitates tracking of interviewing behaviors among interviewers, across studies, and over time. Survey researchers at RTI International have recently developed a monitoring protocol designed to allow quick entry of interviewer behaviors. Our expectation is that this form will provide a complete record of non-standardized behaviors to allow monitors to better assess interviewer knowledge and skills and, therefore, provide more effective feedback. This paper describes the steps in the development of this monitoring protocol and presents monitoring data over time to evaluate the effectiveness of the protocol in improving interviewer behavior.

The authors are grateful to the following call center staff for their contributions to this effort: Richard Heman-Ackah, Rusty Galloway, Sharon Ugochukwu, and all of the monitors who tested versions of the monitoring protocol and otherwise contributed to this effort. We are also very grateful to Lori Madrigal for her considerable efforts to devise (and revise) the electronic monitoring form.
1. Introduction

In order to achieve, and maintain, data quality, telephone surveys must collect data using methods that produce reliable answers to the survey questions. To ensure reliability of survey data, most telephone interviewing seeks to standardize interviewing procedures. Standardized interviewing techniques generally include procedures for reading questions and instructions, probing inadequate answers, recording respondents’ answers, and interacting with respondents.

Monitoring interviewers is one of the most important tools for maintaining data quality in telephone surveys and is generally standard practice in telephone interviewing facilities. In theory, monitoring can play a major role in ensuring that interviewers follow standardized interviewing procedures and other required interviewing practices. Monitoring protocols function to document salient interviewer behaviors and provide interviewers constructive feedback to reinforce prescribed procedures.

Nonetheless, implementing an effective monitoring protocol is not without challenges. Over ten years ago, Couper, Holland, and Groves, (1992) noted that monitoring protocols often (1) tend to follow unsystematic and subjective procedures and (2) include only general impressions of telephone interactions, rather than objective measures of behavior. Although survey organizations have likely made some improvements to monitoring protocols over the past ten years, these observations highlight ongoing challenges for effective monitoring of telephone interviewing.

In this paper, we describe an effort to improve a monitoring protocol based on standardized interviewing principles. We discuss the role of standardized interviewing principles in guiding monitoring protocols, identify some of the main challenges of implementing monitoring protocols, describe key features of the revised monitoring protocol, and examine data on the effectiveness of the revised protocol. Based on these experiences, our conclusions focus on further efforts to systematize monitoring protocols in order to improve our knowledge of effective monitoring practices.

2. Standardized Interviewing Principles and Monitoring Protocols

Adherence to standardized interviewing principles has been one of the guiding forces of survey interviewing over the past few decades. Fowler and Mangione (1990, p. 35) arguably provided the clearest statement of standardized interviewing techniques:

1. Read questions as written.
2. Probe inadequate answers non-directively
3. Record answers without discretion.
4. Be interpersonally nonjudgmental regarding substance of answers.
As Fowler and Mangione (1990) also noted, interviewers often face serious obstacles in trying to follow standardized techniques. Two prime examples they cite are interviewing situations where (1) the goal of accuracy seems to conflict with the goal of standardization or (2) the goal of maintaining rapport appears to conflict with the goal of standardization. Although these authors note that inadequate survey instruments often contribute to these problems, even well-prepared instruments can produce situations where interviewers have difficulty following standardized interviewing techniques. Our approach recognizes that deviations by interviewers from standardized procedures generally increases interviewer variance, and therefore contributes to variable portion of measurement error (Collins, 1980; Dykema, Lepkowski, and Blixt, 1997). Reducing interviewer variability is a key goal of the standardized interviewing approach.

Monitoring telephone interviewers is arguably the most important tool for ensuring that interviewers follow standardized interviewing procedures. Effective monitoring should be able to identify instances where interviewers do not follow standardized techniques so that monitors can work with interviewers to resolve this problem. A key to monitoring, as Cannell and Oksenberg (1988) explain, is the ability of monitors to provide effective feedback to interviewers on the acceptability of their performance. In terms of standardized interviewing, monitors can only provide effective feedback if they can determine:

- whether interviewers understand the principles of standardized interviewing,
- whether interviewers understand the behaviors required to maintain principles of standardized interviewing,
- whether interviewers have the skills to apply the standardized interviewing principles.

Effective monitoring is therefore dependent upon monitors having the tools and data to adequately assess interviewers’ knowledge and skills with respect to standardized interviewing techniques, in order to be able to provide effective feedback.

3. Challenges of Implementing Effective Monitoring Protocols

Monitoring procedures often fail to collect information in a way that facilitates effective feedback to interviewers. Couper, et al, (1992, p. 64) note that systematic monitoring procedures must include the use of forms to facilitate the objective evaluation of interviewer behavior. If a monitoring protocol only collects qualitative ratings of

---

1 Concerns about standardizing interviewing procedures creating conflicts with accuracy or rapport have led an increasing number of researchers to examine alternatives to relying upon standardized interviewing techniques (Conrad and Schober, 2000; Schober and Conrad, 1997). See Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau (2004) and Maynard, Houtkoop-Steenstra, Schaeffer, and van der Zouwen (2002) for recent discussions of this ongoing debate.
interviewer behaviors, monitors are not well-equipped to directly assess interviewers’ knowledge and skills and provide appropriate feedback.

One important challenge of developing and effective monitoring protocol is to determine what data to record; that is, what should be measured and how should it be measured. As Couper, et al (1992, p. 68) noted monitoring data can vary by (1) the unit of measurement, from an individual interviewer utterance to an entire interview, (2) whether the data are objective or subjective, and (3) whether the data include only interviewer behaviors or both interviewer and respondent behavior. In order for a monitoring protocol to provide data that are most useful, behaviors should ideally be recorded at the question level, using objective measures. For example, useful data would be the specific questions where specific types of non-standardized behaviors were observed. Determining whether to focus just on interviewer behaviors or both interviewer and respondent behaviors should probably be considered in light of the purpose of the monitoring protocol. If the primary purpose is to improve standardization of interviewing procedures, focusing solely on interviewer behaviors is likely sufficient. (Couper, et al, 1992). Recording both interviewer and respondent behaviors can be useful to provide greater context to interviewing behaviors, although focusing on both sets of behaviors could make the monitoring protocol too demanding.

A second important limitation of many monitoring protocols is the difficulty for monitors to record specific behaviors during actual interactions or interviews. This is especially true when the interaction is moving at a fast pace. Without an accurate and complete list of the number and type of behaviors observed, it is difficult for monitors to be precise in the feedback they provide to interviewers. This shortcoming also limits survey organizations from using monitoring data to accurately track interviewer behaviors among interviewers, across studies, and over time. Ideally, a monitoring form will allow monitors to quickly record data (with brief notes) during the course of an interview, so that all important behaviors can be included in the data.

4. Improving an Existing Monitoring Protocol

Objectives. Our efforts to systematize our own monitoring protocol focused on the development of a form that provides objective data.2 We followed three objectives in improving our monitoring form:

1. being able to objectively record interviewer behaviors relevant to standardized interviewing procedures,
2. being able to accurately record specific survey questions when relevant interviewer behaviors are observed,

3. providing precise, constructive feedback to interviewers based on accurate, objective data.

A central tenet of our approach is that these three issues are interrelated factors that ultimately determine the effectiveness of monitoring protocols. Monitors cannot provide effective feedback to interviewers if they are not able to identify the frequency and significance of specific interviewing behaviors. Furthermore, monitors cannot easily address specific interviewing skills without placing their observations in the context of specific questions. Following Cannell and Oksenberg’s approach (1988), our expectation was that meeting these objectives would allow monitors to (1) determine interviewers’ level of understanding of standardized interviewing techniques, (2) assess their skill level with respect to these procedures, and (3) provide more precise feedback to interviewers, both positive and negative, to address any improve their skills.

**Procedures:** To develop and implement the revised monitoring protocol, we followed several important steps, a few of which we repeated as necessary:

- **Initial development of the format and content for the revised form.** Our first step was to identify the different interviewer skill areas and specific behaviors to be measured in the revised protocol. We convened an experienced group of “resident experts” to discuss what non-standardized behaviors the form should measure and how it should measure these behaviors. This brainstorming session provided the basis of our first draft of the new form, which we continued to revise and improve at each stage in the process.

- **Group discussions with monitors, supervisors, and interviewers.** To engage all major stakeholders in the monitoring protocol, we first held discussion groups with monitors and supervisors. These discussions were informal and focused on some of the limitations of the current monitoring form and desired outcomes of the revised protocol. We also extended these discussions to include additional call center managers and interviewers to get their perspective on improving the monitoring protocol. As we made significant revisions to the form and new monitors were hired, we held further sessions to continue to receive feedback on the protocol.

- **Usability testing of a draft version of the form.** In order to test our first draft of the form in realistic conditions, we used portable usability testing equipment to record monitors as they tried out the new form during actual monitoring sessions. These sessions provided valuable information about the interface of the draft form and the monitor’s job, and led to some significant revisions in the protocol.
• **Revisions of the form based on monitor input and testing results.** One of the key themes that emerged from our discussions and testing sessions with monitors was the challenge of using the monitoring software, CATI software, and a monitoring form at the same time. This challenge led us to find ways to improve the usability of the form so that monitors could focus on their task more easily. For example, we found that monitors were having a difficult time making notes when using both the previous and draft version of the new form. We therefore modified the notes page to function more like a standard word processing program. By simply clicking on the notes section, the new form changes to a word processing environment. This feature allowed monitors to more easily record important notes without distracting them from the interviewer’s work.

• **Training monitors on using the new form.** Since the revised monitoring protocol represented a significant departure from the previous form, we scheduled sessions for monitors to discuss the goals of the new protocol, (especially the focus on standardized interviewing), review each section and feature of the new form, and learn how to navigate and use the new form. We created an instruction manual to provide guidance for the training sessions and act as an ongoing resource for monitors.

• **Follow-up with call center staff.** Once the revised protocol was fully implemented in the call center, we continued to solicit feedback from monitoring staff on the usability of the form and the results of monitoring sessions. This feedback suggested further adjustments the form, especially calibrating the scoring to more accurately reflect interviewer performance.

At each stage of development, we focused on achieving our primary goal of facilitating easy entry of objective indicators of interviewer behaviors during monitoring sessions.

During the development of the new form, we benefited input from a variety of staff to inform our approach. One conclusion we reached early in the process was that we needed to create the form in a program that would provide flexibility in recording, calculating, and presenting the results of monitoring sessions. We decided to create the form in Excel, which provides automatic tallying of monitors’ entries and calculation of scores and offers other features compatible with our goals. Another important strategy we adopted fairly early was to maintain consistency in how specific items in each interviewing skill area are presented. For example, in the interviewing skills section, we chose to present all items as “negatively” scaled. That is, this section of the form only specifies that monitors record behaviors that violate standardized interviewing principles,
not those that conform to, or help reinforce, standardized procedures.\(^3\) The clear advantage of this approach is that, in most interviewing situations, it is nearly impossible for a monitor to record every positive standardized behavior that a well-trained interviewer follows. Our approach focuses the attention of monitors and interviewers on specific behaviors that need to be addressed in order to maintain standardization.

5. Key Features of the Revised Monitoring Protocol

The most important feature of the revised monitoring protocol is facilitating quick entry of non-standardized interviewing behaviors, such as adding unscripted text to questions or using a leading probe. The revised monitoring form allows non-standardized interviewing behaviors to be more easily identified by monitors during the interaction with a respondent. Furthermore, the Excel form organizes the monitoring session data and handles administrative tasks (such as tallying scores) so that monitors can focus their attention on interviewers’ behaviors.

The monitoring form is divided into six sections, which are organized as six different Excel sheets within the form. The six sections are:

1. Call Management Skills
2. Introduction / Obtaining Cooperation
3. Interviewing Skills
4. Presentation Skills
5. Monitor Notes and Comments
6. Overall Score and Comments

Since Excel provides a tab for each sheet at the bottom of the form, moving from one section to the next is easy. Since monitoring sessions vary considerably in what aspects of telephone data collection are observed, from unanswered calls to complete interviews, the specific sections of the form that a monitor has to complete can be matched to the activities observed during the session.

Since the focus of our research is on standardized interviewing techniques, our primary concern is the Interviewing Skills section. This section is divided into five skill areas – Speech Characteristics, Reading Skills, Probing Skills, Professional Handling of Interview Situation, and CATI Skills. Descriptions of each of these five skill areas and how they relate to standardized interviewing principles are provided in Figure 1. These descriptions are provided in an instructional manual for monitors.

---

\(^3\) This strategy met considerable resistance from monitors who were concerned about the “negative” focus of the form. That is, the form does not directly record positive, standardized behaviors, only those behaviors that violate standardized interviewing procedures. Interestingly, Couper, et al (1992), drew similar reactions from monitoring staff in implementing a new monitoring form.
1. **Speech Characteristics**  
The purpose of this skill area is to provide information on the way the interviewer controls her voice and responds to the respondent’s need for clarity of pronunciation. These behaviors relate to standardization of the presentation of survey questions. The monitor should record any instances where the interviewer does not articulate words clearly or pronounce words correctly. No items in this skill are critical items.

2. **Reading Skills**  
The purpose of this skill area is to measure how well the interviewer reads the question text verbatim and follows instructions on what should be read on each question. The monitor should record times when the interviewer adds or omits word from the question text, or incorrectly reads unspoken response categories. These behaviors also relate to standardization of the presentation of survey questions and can directly affect the meaning of items. Two items (“Omitted major word” and “Failed to read an entire question or instruction”) are critical items.

3. **Probing Skills**  
This is an especially important section, because it is critical to maintaining non-biased, standardized interviewing when survey participants do not initially provide adequate answers. Questions that require probing make it more difficult for interviewers to maintain standardized data collection. The form allows the monitor to count the number of times the interviewer uses a leading probe or ignores a situation when a probe would be appropriate. None of the items in this skill area is a critical item.

4. **Professional Handling of Interview Situation**  
This section deals with treating the respondent with respect, maintaining appropriate rapport with respondent, and conveying knowledge of the study to the respondent as needed. These items relate to maintaining a neutral, nonjudgmental stance toward the substance of participants’ answers and comments.

5. **CATI Skills**  
This section allows the monitor to record times when the interviewer does not use CATI or the keyboard correctly, such as incorrectly coding responses or not keying verbatim answers fully. These items relate to standardization of the recording of respondents’ answers. Two items in this skill area are critical items.
Some key features of this section provide a number of improvements over the previous protocol to facilitate more accurate and complete monitoring data and more effective feedback to interviewers:

- The form provides multiple spaces for any non-standardized interviewing behaviors that are observed, so that monitors have complete and precise information to provide interviewers during the feedback session. Any question number entered in a cell (e.g., A16, BXXTF, or income) is treated by the form as 1 incidence of a non-standardized behavior.

- The form includes specific behaviors under each area, which provide the basis for the overall score for each section. As a result, the form provides detailed information on specific interviewing behaviors as well as a summary rating of each interviewing skill area.

- The form highlights “critical” non-standardized behaviors that interviewers must avoid in order to obtain a passing score for the monitoring session. Attention to how interviewers are doing on these items emphasizes adherence to the underlying standardized interviewing principles.

Furthermore, like other sections of the form, entries for each specific behavior are tallied automatically by the form to produce the skill area scores. These data facilitate tracking of interviewing behaviors among interviewers, across studies, and over time, so that survey managers can identify broader trends in interviewers’ behavior.

6. Evaluating the Effectiveness of the Revised Monitoring Protocol

Evaluating the success or failure of any programmatic change that occurs in organizational setting is often a daunting task. We considered, and mostly rejected, different kinds of data we could examine to evaluate the effectiveness of the revised monitoring protocol. The two main challenges to identifying data to assess the impact of the revised form are the validity of the data and the timing of the data. As one approach, we considered using overall monitoring scores in our analysis. An important limitation of these data is that the overall scoring criteria and some of the skill area scoring criteria changed during the first year of implementation. Hence, some portion of the change in the monitoring scores can be attributed to changes in the scoring procedures, rather than any changes in interviewer behavior. Recalibrating the scores to reflect these changes is not really a viable option to produce comparable data across time. Likewise, necessary adjustments to the revised monitoring protocol over the course of implementation period made it difficult, if not impossible, to fix two (or more) clear time frames that could legitimately be viewed as the pre-implementation and post-implementation periods. These limitations are the direct consequence of conducting research on procedural
changes that needed to meet the operational needs of a survey organization. This research is based on the actual implementation of a revised monitoring protocol in a call center, rather than experimental conditions in a survey laboratory.

Given the constraints on the data available to evaluate the effectiveness of the revised monitoring protocol, we concluded that the best approach was to look at the monitoring data over multiple time periods that are not associated with specific stages of implementation. Organized in this way, the data will provide us a simple trend in the monitoring data that encompasses the different sources of variation over the entire implementation period. For this analysis, we simply organize the monitoring data in standard calendar-year quarters. These data will allow us to simply evaluate the trend in monitoring results over the first full year of implementation. We also recognize that we are combining data across different sets of studies with different characteristics each quarter. Variation in the complexity and quality of the survey instruments could obviously affect the ability of interviewers to adhere to standardized interviewing principles. Our approach assumes (1) the same interviewing standards are applied across studies for each quarter and (2) variation among survey instruments produce a comparable set of studies for each quarter.

Second, our focus on standardized interviewing led us to limit the present analysis to interviewer behaviors during administration of an interview. Although the monitoring form tracks a comprehensive set of interviewer behaviors across the different types of interactions and situations that interviewers encounter, the greatest impact of standardized interviewing techniques is observed in interview sessions. For this reason, we focus our evaluation on the third section of the monitoring form, which addresses interviewing skills. We measure two outcomes: (1) the mean number of non-standardized behaviors for each interviewing skill item across all sampled forms and (2) the mean skill area score for each of the five interviewing skill errors. These data will provide a detailed picture of the general trend in specific interviewer behaviors and interviewing skills over the first year of implementation.

Our primary expectation is that the trend in the monitoring data will reveal a general decline in the number of non-standardized behaviors reported in monitoring sessions and, therefore, a general increase in the score for each skill area. This expectation reflects our general hypothesis that the more precise recording of non-standardized behaviors will result in more effective feedback to interviewers. We anticipate that this feedback will clarify specific non-standardized behaviors interviewers made and the context in which they made them. This expected consequence of the revised protocol will allow interviewers to better understand how and why they engaged in these behaviors. Following Cannell and Oksenberg’s (1988) arguments, we also anticipate interviewers will be more successful in bringing their behavior in line with standardized techniques and thereby decrease non-standardized incidences.
In order to conduct this analysis, we compiled monitoring data from various surveys across the four quarters of the 2005 calendar year using a multi-stage sampling approach. We first identified the studies conducting data collection in each of the four quarters. From the set of studies available for each quarter, we then sampled forms across all interviewers on each study. As the final stage, we then sampled monitoring forms for each interviewer that included observations of a complete or partial interview. Since monitoring forms are completed for sessions where no interviewing was observed, we had to sample additional forms for each interviewer to account for the significant number of “ineligible” forms without any data in the Interviewing Skills section from the analysis. This analysis will provide preliminary evidence on the impact of the revised monitoring protocol on adherence to standardized interviewing techniques during the first year of implementing the new protocol.

7. Results

Table 1 presents quarterly data from sampled monitoring forms. The quarterly samples were 33, 49, 45, and 51 monitoring forms. Variation in the sample size across quarters was the result of variation in the number of telephone surveys being conducted and, to a lesser degree, variation in the proportion of sampled forms that included observation of a complete or partial interview.

For each of the non-standardized behaviors tracked in the interviewing skills section, we compiled the quarterly average number of times each specific behavior was observed in the sampled monitoring sessions. Behaviors appearing in the table in **bold italic** text are “critical” items that automatically produce an unacceptable score (“did not meet expectations) for the entire skill area when observed in a monitoring session. The data on non-standardized behaviors are directly comparable across all four quarters, since the same monitoring instructions and procedures applied to all quarters.

Looking at the specific behaviors monitored, the most striking result was the rarity of non-standardized behaviors observed in the sampled monitoring forms. For five of the 18 specific behaviors monitored, no instances were recorded among the sampled forms. For another seven behaviors, instances were only observed in only one of the four quarters. As a result, the average rates for non-standardized behaviors were all quite low. The most common behaviors observed across all four quarters were **failing to read an entire question or instruction**, **ignoring unclear responses**, and **using non-neutral recognition**. Three other specific behaviors were observed (at least once) in multiple quarters: **reading response categories incorrectly or reading uppercase text**, **providing clarification inconsistent with project-specific instructions**, and **using a leading probe**.

The relatively rare incidence rate of non-standardized behaviors makes it difficult to discern any clear trends in incidences of non-standardized interviewer behaviors.
<table>
<thead>
<tr>
<th>Interviewing Skills</th>
<th>Q1 2005 Average Behaviors</th>
<th>Q2 2005 Average Behaviors</th>
<th>Q3 2005 Average Behaviors</th>
<th>Q4 2005 Average Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not articulate words clearly</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Incorrectly pronounced a major word</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Reading Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Added unscripted words to the question text</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Omitted major word (such as noun or verb)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Omitted minor word (such as &quot;the, of, a, and&quot;)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Read response categories incorrectly or read uppercase text</td>
<td>0.15</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Failed to read an entire question or instruction</strong></td>
<td>0.09</td>
<td>0.06</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Probing Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ignored unclear responses</td>
<td>0.06</td>
<td>0.02</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Provided clarification inconsistent with project-specific instructions</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Did not probe &quot;don't know&quot; or &quot;refused&quot; answer</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Used a leading probe</td>
<td>0.06</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Professional Handling of Interview Situation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used non-neutral recognition (e.g., “I hear ya”)</td>
<td>0.15</td>
<td>0.00</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Provided incorrect response to project question</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Allowed respondent to stray from task</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>CATI Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coded responses incorrectly</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Did not spell verbatim answers correctly</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Did not fully key verbatim answers</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Did not verify spelling of contact information</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Number of Monitoring Forms</strong></td>
<td>33</td>
<td>49</td>
<td>45</td>
<td>51</td>
</tr>
</tbody>
</table>
Three behaviors with relatively high incidence rates in the first or second quarters (reading response categories incorrectly, failing to read an entire survey item, and using non-neutral recognition) appeared to decline somewhat over the course of the year. Two other behaviors (ignoring unclear responses and using a leading probe) also appeared to decline during the year, although incidences of these behaviors were quite low overall. Other behaviors were relatively less common and followed no clear pattern over the course of the year. Although the data on non-standardized behaviors show, if anything, a slight decline over the year, the rarity of these behaviors makes it difficult to discern whether these patterns reflect meaningful variation.

Additional data from the monitoring forms, which we do not present in Table 1, are the five skill area scores. Each of the skill area scores had three categories in the first two quarters: 0 = exceeded expectations, 1 = met expectations, and 2 = did not meet expectations. Based on feedback from monitors and call center managers, we modified the scoring categories for three of the five skill areas (Speech Characteristics, Reading Skills, and CATI Skills) at the beginning of quarter 3 so that only scores of 1 or 2 are allowed. As a result, the skill area scores are comparable between the first and second quarters and between the third and fourth quarters, but not across the first two and last two quarters of the year. Still, to see how the skill area scores looked over the year, we calculated the average scores for each of the five interviewing skill areas for each quarter. Overall, the scores for the five skill areas reflect the relatively low incidences of non-standardized behaviors and, as a result, exhibit even less variation than the behaviors. In the first two quarters of 2005, the average scores were close to zero or, for some scores, actually zero. Similarly, in the third and fourth quarters, skill area scores were close to 0 for Probing Skills and Professional Handling of the Interview Situation and exactly 1.00 for the other three skill areas. The consistent scores of 1.00 for three of the skill areas indicate ratings for these areas were uniformly “met expectations” on all sampled monitoring forms. The variation in skill area scores did not appear to demonstrate any clear trends. From quarter 1 to quarter 2, scores for two skill areas increased, scores for one area decreased, and scores for the other two areas remained the same. Similarly, from quarter 3 to quarter 4, scores for four of the five areas stayed the same and only one area (Probing Skills) showed a decline in the score.

8. Discussion

This paper sought to contribute to the rather sparse recent literature on monitoring telephone interviewing. The most important contribution of this research was to further systematize monitoring procedures by developing a monitoring form that collects more objective data on interviewer behaviors. This approach facilitates (1) quantifying inappropriate interviewer behaviors more precisely and (2) providing more instructive feedback to interviewers about their behaviors.
Data from the first year of implementation of the new monitoring protocol provided rather limited evidence on the effectiveness of the new form. Incidences of non-standardized behaviors were relatively rare overall, making it difficult to discern significant trends in interviewer behavior. The limited patterns observed did indicate decreases over the year in a few interviewer behaviors, including (1) reading response categories incorrectly, (2) failing to read an entire survey item, and (3) using non-neutral recognition. Although the declines in these behaviors were quite modest, they provide some evidence that these inappropriate behaviors were being reduced over the first year of implementation of the form. Furthermore, no interviewer behaviors showed a clear increase over the year, an important outcome when significant numbers of new telephone interviewers are being hired.

Some comparisons between our data and similar data from Couper, et al (1992) are instructive. Their data are not directly comparable to ours, since they calculate the percentage of errors observed for all questions monitored rather than the average number of behaviors observed in monitoring sessions. On the other hand, the data from both studies reflect measures of the rate at which incidence of specific interviewer behaviors occur. Similar to our results, Couper, et al (1992) found relatively low incidences (or “error rates”) for most monitored interviewer behaviors, including 0 instances for a few behaviors. At the same time, they did report somewhat higher incidences for most behaviors.

Another important comparison with Couper, et al (1992) involves the relative incident rates of different interviewer behaviors. Although the set of behaviors observed in our study and theirs were not identical, a number of the same behaviors were monitored in both research efforts. One striking difference between the two studies is that Couper, at al (1992) observed greater incidences of minor wording changes than in our study. We observed relatively few incidences of omitting minor words or adding unscripted words to questions over the four quarters. Under Probing Skills, the two studies appeared to observe similar incidences of inappropriate probing, directive probing, and failing to probe when needed. The data on probing are important to note, since questions that require probing have a high probability of producing variability in interviewer behavior. For this reason, probing skills are often considered to be “higher level” interviewer skills that require considerable experience, training, and monitoring.

Some important limitations of our research leave our results and conclusions tentative. We were only able to sample a limited number of forms for each time period. A somewhat larger sample, or repeated analysis across additional quarters, might provide clearer trends in interviewer behaviors. Furthermore, our analysis of successive quarters includes multiple sources of variation that cannot be untangled. For example, the number of non-standardized interviewer behaviors is likely to increase, at least temporarily, when significant numbers of new interviewers are added to the staff or when new studies
present particularly challenging instruments. Couper, et al (1992) found some important differences in comparing new interviewers to more experienced interviewers. Our analysis focused on the overall trend in monitoring data, without specifically controlling for these kinds of variation over time. Still, the overall trends are clearly an important indicator of the impact of the revised protocol across different sets of interviewers and different studies.

Our experiences also suggest a couple further avenues for research and practice in improving monitoring protocols to maintain standardized interviewing. First, as Couper, et al (1992) recommended, systematizing monitoring procedures also requires attention to how interviewers are sampled for monitoring. The relatively low incidence of non-standardized behaviors might largely be the result of strong performance by well-trained interviewers, but could also be partly due to not sampling sufficient numbers or durations of telephone interviews. Using equivalent probability sampling procedures clearly provides the most statistically sound representation of interviewers’ behavior. One practical concern about following such sampling procedures is that managers and monitors often want to monitor larger segments of interviews and devote additional effort to monitoring certain interviewers (such as those who are newly-hired or have noted performance issues). For these reasons, monitoring sampling strategies must be carefully considered. A balance between probability and purposive sampling is likely to be the most effective model. Developing models to address both the systematic and practical demands of monitoring would help to further systematize monitoring procedures.

Second, systematizing monitoring observations at the question level can provide information both on interviewer performance and problem questions. While our focus has been on interviewer behaviors, the information on specific survey items can be related to interviewer performance. Understanding the kinds of questions that pose the greatest challenge to standardized interviewing techniques can provide survey managers with valuable information for training and supervising interviewers. For example, researchers have shown that questions which require probing generally produce greater variability in interviewer behavior (Collins, 1980; Fowler and Mangione, 1990). Systematic monitoring data can be used to identify specific types of questions that cause the most difficult for interviewers to administer in a standardized way, so that training and supervision procedures can be devised to address administration of these items.

Our research represents an attempt to further systematize a monitoring protocol based on standardized interviewing principles and the importance of effective feedback to monitoring telephone interviewers. We extended these principles by developing a monitoring protocol to capture complete, accurate, and objective data in monitoring sessions. These efforts remind survey researchers of the importance of capturing appropriate information to understand how interviewers affect survey data.
References


