Developing a Common, Mode-independent, Approach for Evaluating Interview Quality and Interviewer Performance

Howard Speizer, Susan Kinsey, Richard Heman-Ackah, M. Rita Thissen

RTI International

Abstract

RTI is developing an interview quality monitoring evaluation system based on audio recordings of the survey interview to be used across all interviewer-administered survey modes. In-person and telephone interviewers are evaluated by similar quality metrics and their performance is tracked across projects. The interviewer evaluation, feedback, and coaching processes have been standardized, helping make the quality monitoring process more efficient. The system provides an easy-to-navigate front-end to all recorded survey interactions that quality supervisors, project staff, and clients can use. This paper examines the design of this new system, including the strategies used to standardize evaluation metrics across survey modes, store and analyze digital files, and develop interviewer- and survey item-level quality evaluation databases. We discuss the system's fit across a wide spectrum of interview quality monitoring processes. The paper concludes with planned future uses and enhancements to the system.

Key Words: quality assurance, audio recording; computer audio-recorded interviewing (CARI); quality monitoring; interviewer performance evaluation; standardized quality monitoring system; Quality Evaluation System (QUEST); telephone interview; field interview; in-person interview

1. Introduction

Interview quality is an important and closely monitored component of overall survey data quality with considerable attention paid to interviewer performance. We implement interview quality control protocols that include extensive interviewer training and practice drills, monitoring of the work being performed, and direct feedback to improve performance. Whether we are monitoring in-person (field) or telephone interviewers, our goal is to authenticate the interview, verify that the interviewer followed the data collection protocol, and ensure that he/she correctly administered the survey instrument. Direct feedback after quality monitoring is designed to improve interviewer performance and identify systematic or special-cause problems that require revisions to protocols, training programs, or instruments.

Historically, RTI has used several mechanisms for monitoring interview quality including silent monitoring, telephone verification interviews, direct (in-person) observations, validations, and data reviews, to name a few. In 1999, RTI pioneered the integration of computer audio-recorded interview (CARI) technology with computer-assisted personal interviewing (CAPI) software for silent recording of the interviewer-respondent interactions that occur during administration of an in-person interview (Biemer et al., 2000). CARI enables direct and timely assessment of the quality of an interviewer's performance in the field without incurring the high costs that can be associated with direct interview observations.

Although CARI has proven to be an effective tool for evaluating field interviewer performance, RTI has only recently attempted to use audio recordings to evaluate telephone interviewer performance. Telephone surveys, like field surveys, can benefit from the collection of audio recordings for quality review and interviewer performance assessment. Today, most off-the-shelf digital CATI systems offer digital recording as a standard feature and digital audio visual recording is a mature technology for monitoring telephone interactions.

Because the monitoring systems for in-person and telephone interviews evolved independently at RTI, performance criteria, scoring systems, and feedback methods diverged. In addition, study teams were allowed to develop customized monitoring protocols and define new metrics for evaluating interviewer performance, which at times conflicted with existing standards. In response, a team of senior research staff[1] was commissioned in fall 2008 to develop a cohesive, standardized quality monitoring strategy and a system and set of processes to support that strategy for all RTI surveys. At the core of this strategy is the intent to maximize the use, and impact, of CARI technology for monitoring both telephone and field surveys. In this paper we describe the overall quality monitoring strategy and RTI's new standardized quality monitoring system—QUEST (Quality Evaluation System)—developed for monitoring interviewer performance.

2. Objectives of Standardized Quality Monitoring System

RTI's QUEST system is intended to meet several critical objectives, including (1) standardization of monitoring protocols, metrics, and feedback mechanisms; (2) collection of trend data to evaluate interviewer and survey item-level performance within and across projects; (3) increased efficiency of monitoring operations; (4) increased use of audio recording (CARI) to evaluate and improve interviewer performance; and (5) improved access to interview data by clients. Each objective is discussed below.

Standardization of Monitoring Protocols, Metrics, and Feedback Mechanisms

Given the use of multiple quality monitoring systems and tools at RTI, there has been variability in the skills and behaviors examined, the performance metrics monitored and tracked, and in the

[1] The authors gratefully acknowledge the contributions of the other members of the RTI QUEST development team: Mary Allen, Melissa Cominole, Doug Currivan, Orin Day, David Foster, Courtney Gainey, Kristen Miller, Mai Nguyen, and Sridevi Sattaluri.

nature and timing of the feedback delivered to interviewers. Thus, important objectives for QUEST developers have been identifying those critical interviewer skills and behaviors requiring evaluation in telephone and/or field surveys (i.e., probing, question administration, case management); designing an evaluation form and scoring rubric that supports a fair evaluation of these skills and behaviors; and developing a standard, mode-independent performance tracking and feedback process that drives continuous quality improvement.

Collection of Trend Data to Evaluate Interviewer and Survey Item-level Performance
The QUEST system has been designed to help research staff evaluate trends across projects and data collection modes. The objective is to use historic data for longitudinal assessments of interviewer performance against specific quality dimensions (for example, probing), and support more rigorous examinations of performance data by interview mode and other important factors. Standardized evaluation metrics and reports make it possible to measure improvement or "drift" in skills over time and identify common problems in field and/or telephone data collections to inform interviewer training programs.

Increased Efficiency of Monitoring Operations
Live monitoring in the call center environment relies on "snapshots" of an interviewer's performance, that is, time-slices that provide monitors a picture of what the interviewer is doing at the time they tune in via silent monitoring. While this allows monitors to hear a wide variety of interviewer-respondent interactions, inefficiencies are inherent in this process because monitors spend time plugging into nonproductive calls (e.g., busy signals) and searching for productive calls that can be monitored. With recorded interactions, however, monitors can efficiently target specific interview sessions and calls, avoiding the inefficiencies inherent in the silent monitoring process, while still providing timely, constructive performance feedback to interviewers.

Increased Use of Audio Recording (CARI) to Evaluate and Improve Interviewer Performance
Beyond efficiency, the RTI team sought to improve interviewer performance by leveraging opportunities provided by digital recording, some of which include:

- Improved interviewer feedback—constructive feedback that is supported by a recording of the interview has more impact than feedback provided from direct observation (or live monitoring). By listening to his/her voice, the interviewer can more easily improve voice quality and tone and understand more fully any interview administration mistakes that were made.
- Survey item performance—audio recordings can be categorized and evaluated by question, or question type, providing important insights into survey instrument performance. This is particularly valuable during a pretest or early in the field period.
- Inter-rater reliability—recordings can be used to train monitors on evaluation protocols, thereby ensuring consistency in interviewer feedback and minimal rater variability;

- Ability to verify up to 100% of interviews—traditional monitoring protocols target a sample (5–15%) of interactions. With digital recording all interviews can be recorded and different samples can be selected for different quality assurance purposes.
- Response latency—recordings can be used to measure and evaluate respondent cognitive difficulty in answering complex, challenging, or sensitive questions. Video recordings may also be used to evaluate how well interviewers navigate complex screen layouts and code respondent responses.

Improved Access to Interview Data by Clients
QUEST is intended to give internal and external clients increased access to interview data via recorded interviewer-respondent interactions, thereby minimizing technology barriers and reducing the need for more costly direct observations of interviews. The goal is to provide our clients with greater opportunity and a more efficient, less burdensome means to hear and see how their survey items are administered and how respondents receive them.

3. Practical Design Considerations

In arriving at a cost-effective, yet flexible strategy for quality monitoring that achieves the desired objectives, RTI developers have considered several important practical design elements.

Recording and Monitoring Standards
Operational standards have been established for QUEST quality monitoring at RTI. These include recording and monitoring rates designed to provide sufficient review of performance and adequate feedback to interviewers. These standards can be revised for special quality monitoring circumstances and permanently modified when necessary. The current recommended standards include the following:

- The first two interviews that (field and telephone) interviewers complete are recorded and monitored.
- A minimum of 10% of each interviewer's completed interviews are selected for monitoring, or more depending on contract requirements, or to investigate quality concerns.
- For purposes of authenticating an interview, it is recommended that a minimum of 3 minutes of audio be reviewed.
- For purposes of interviewer performance feedback, it is recommended that a minimum of 6 minutes of audio be reviewed. This accomplishes the authenticity assessment while providing monitors with additional information upon which to rate an interviewer's performance. Monitoring in 6-minute sessions is also consistent with current "live" monitoring standards practiced in the call center.
- Projects are encouraged to vary what they record during the interview so that interviewers don't know which items will be reviewed and to provide monitors with a mix of interactions to evaluate.

- For field projects, survey managers are also encouraged to record and review interview protocols such as gaining informed consent or administering respondent incentives in part to reduce the need for telephone verifications or other quality monitoring steps.

Thissen et al. (2008b) reported on a small experiment that suggested three 30-second audio files may be adequate for verification purposes. This work focused primarily on authenticating interviews, however, and projects often review more audio to support interviewer performance evaluation. Additionally, Tarnai (2007) concluded that partial monitoring sessions were the "most effective and efficient monitoring procedures." The recommended standards established by RTI achieve the goal of performance monitoring in an effective yet cost-efficient manner.

QUEST employs an algorithm for sampling the available audio files for quality review with the following capabilities:

- Select the first *n* (default is 2) cases for any new interviewer.
- Select the n^{th} case for each interviewer, according to a person-specific review percentage.
- Apply a default percentage to each interviewer for whom no person-specific review percentage is available.
- Allow certain cases to be excluded from review (such as a situation where a respondent retroactively withdraws permission).
- Allow certain cases to be forced into review (such as a situation known to be problematic).

The QUEST system allows quality supervisors to adjust the percentage of cases sampled for each interviewer. As noted, however, we continue to explore the optimum amount of monitoring required based on the type of survey or interviewer activity, prior data about the interviewer's performance, cost constraints, and more specific quality performance goals.

Performance Evaluation

Figure 1 provides a sample of the performance evaluation blocks developed for QUEST, including the evaluation items within each block and the associated item scoring categories. The interviewer performance evaluation scheme was developed based on best practice protocols already used at RTI. The evaluation items include a set of incorrect behaviors that the quality monitor flags during the review of the interviewer-respondent interaction. A set of positive behaviors (e.g., demonstrated excellence in the handling of refusals) is also included to capture exemplary performance by the interviewer. The evaluation items are grouped by performance dimension, in a modular fashion, to facilitate the inclusion or exclusion of item blocks based on the type of interview being reviewed. For example, two Reading Skills blocks are included in the evaluation form—one for projects employing conventional, or verbatim, interviewing methods, and one for projects utilizing conversational interviewing techniques. For a given project, only one of these two evaluation blocks is applicable. The modular design also provides greater flexibility to the system by enabling the addition of new item blocks in the future.

Figure 1. Example of QUEST Standardized Performance Evaluation Criteria and Associated Item Scoring Categories

CASE MANAGEMENT SKILLS	ITEM SCORING CATEGORY
Problem sheet missing/inadequate	**NC**
Callback/appt not set/inaccurate	**CR**
Answering/privacy mgr message incorrect	**NC**
Screen navigation error	**NC**
Case coding error	**CR**
Case notes/comments inappropriate	**CR**
INITIAL CONTACT	ITEM SCORING CATEGORY
Identification of self/RTI/sponsor inappropriate	**CR**
Verification of R's information incorrect	**CR**
Response to R's study questions inadequate	**NC**
Refusal aversion techniques inappropriate	**CR**
Confidentiality concerns/voluntary nature improperly addressed	**NC**
Informed consent not verbatim	**CR**
READING SKILLS—CONVENTIONAL INTERVIEWING	ITEM SCORING CATEGORY
Articulation unclear	**NC**
Pronunciation incorrect	**NC**
Unscripted words added to text	**CR**
Major word/phrase omitted	**CR**
Minor word/phrase omitted	**NC**
Response categories not read when required	**CR**
Entire question/instruction omitted	**CR**
EXCEPTIONAL ITEMS	ITEM SCORING CATEGORY
Interview administration superior	**EX**
FAQs used in exceptional way	**EX**
Difficult situation handled well	**EX**
Proper refusal conversion techniques demonstrated	**EX**
Proper response to difficult R demonstrated	**EX**
Proper response to emotional distress demonstrated	**EX**
Notes: DK = Don't Know FAQ = Frequently Asked Question HH = Household Q = Question R = Respondent RE = Refused	**Scoring Key:** CR = Critical error EX = Exceptional behavior NC = Noncritical error SR = Supervisor/technical review

Evaluation items are weighted for criticality. Some behaviors require immediate supervisory attention (falsification, for example); otherwise, the weights are used in the construction of the block and overall session scores. After the monitor rates the interviewer's performance, the system generates a score for each rated block and for the session overall based on the number and criticality of the behaviors observed. Because a recording may not present the monitor with an opportunity to rate the interviewer's performance in each dimension, monitors use a checkbox to indicate that an item block does not apply. Only blocks rated by the reviewer are included in the overall session score. Positive or "Exceptional" behaviors can improve the overall score for the session and are included in feedback to the interviewer. Additionally, monitors can enter both block- and overall session-level comments to provide interviewers with details about positive or negative behaviors observed

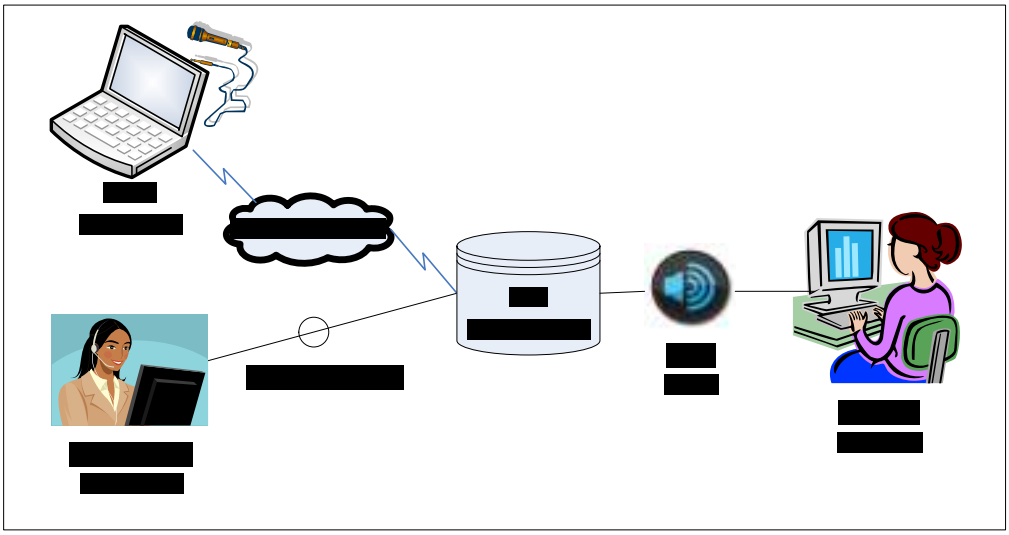
Performance Reporting

Performance reports provide qualitative and quantitative metrics regarding individual interviewers on specific surveys and across surveys, composite overviews of entire surveys, and insight into the quality monitoring process itself. QUEST has been designed to provide projects with comprehensive, standardized reports to track the progress of monitoring activities and to enable performance monitoring at various levels (i.e., interviewer, supervisor, project, overall). Results from the performance evaluation are compiled and feedback provided in summary form for interviewers; this includes block- and overall session-level scores, monitor comments entered in each block or for the overall session, and counts of specific negative (or positive) behaviors observed during the monitoring session. Data are also available to projects through QUEST to support customized analyses and the creation of project-specific reports.

4. Software Design Details

QUEST is designed to support quality monitoring independent of data collection mode. Audio recordings from both telephone and in-person interviews are stored and evaluated similarly through a common user interface. Figure 2 provides a conceptual overview of CARI for telephone and field surveys with audio review through the QUEST system. At the heart is a relational database supporting the QUEST website. Into the database, interconnectivity software writes the location of audio recordings that were collected on field laptops or through the call center's Voice over Internet Protocol (VOIP) system. Other interconnectivity applications populate tables with paradata and response data from field and telephone data collection. The quality monitor opens up the QUEST website, selects a case to work, listens to audio files, and completes the evaluation form. When the monitor indicates the review is complete, the system automatically scores the session and uses the scores to generate reports on demand.

Figure 2. Conceptual View of the QUEST Quality Monitoring System



Recording Interviews

On field laptops, audio recording can be started and stopped by logic embedded within the questionnaire software, creating sound files on the computer's hard drive. The files are compressed for ease of storage and encrypted for security, after which they are returned to a central location along with other forms of response data, such as the case data file. The audio file can be seen as simply another mechanism for capturing information during the interview. Non-interview activities in field surveys usually do not offer an option for recording because they take place without the aid of any electronic device; locating and contacting activities, eligibility screening, gaining cooperation and other interactions generally occur before the interviewer turns on a computer, unless a handheld device is employed at the doorstep.

In the call center, the situation differs from the field, because transmission takes place within a single network. Audio recording may include the entire duration of each phone call if the center employs a digital phone system using VOIP. With VOIP, the digitized voices can be stored by the telephone system as audio files on the server for later review.

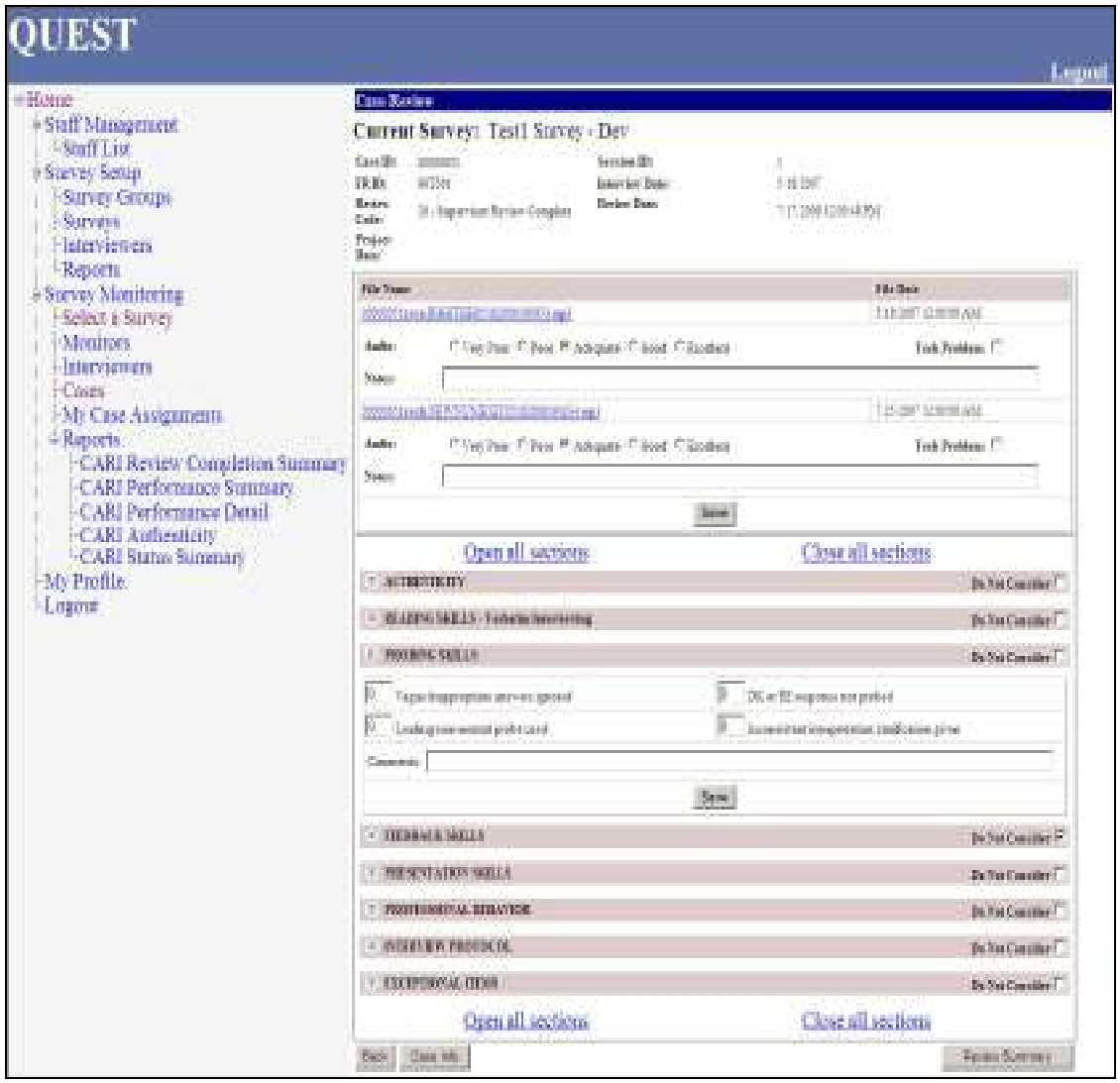
Platform

The QUEST system runs as a Visual Studio.Net Web application hosted on RTI's internal network and uses a SQL Server database. The choice of a Web interface provides access to multiple users at a single time without danger of corrupting the database and without the need for any specialized software on their desktop computers. Use of a relational database facilitates management of the audio recordings and the evaluation data.

User Interface
Within QUEST, the quality monitor is presented with a list of assigned cases and an evaluation page. The contents of the evaluation page are consistent across surveys and simultaneously adaptable to specific surveys through configuration of the evaluation blocks.

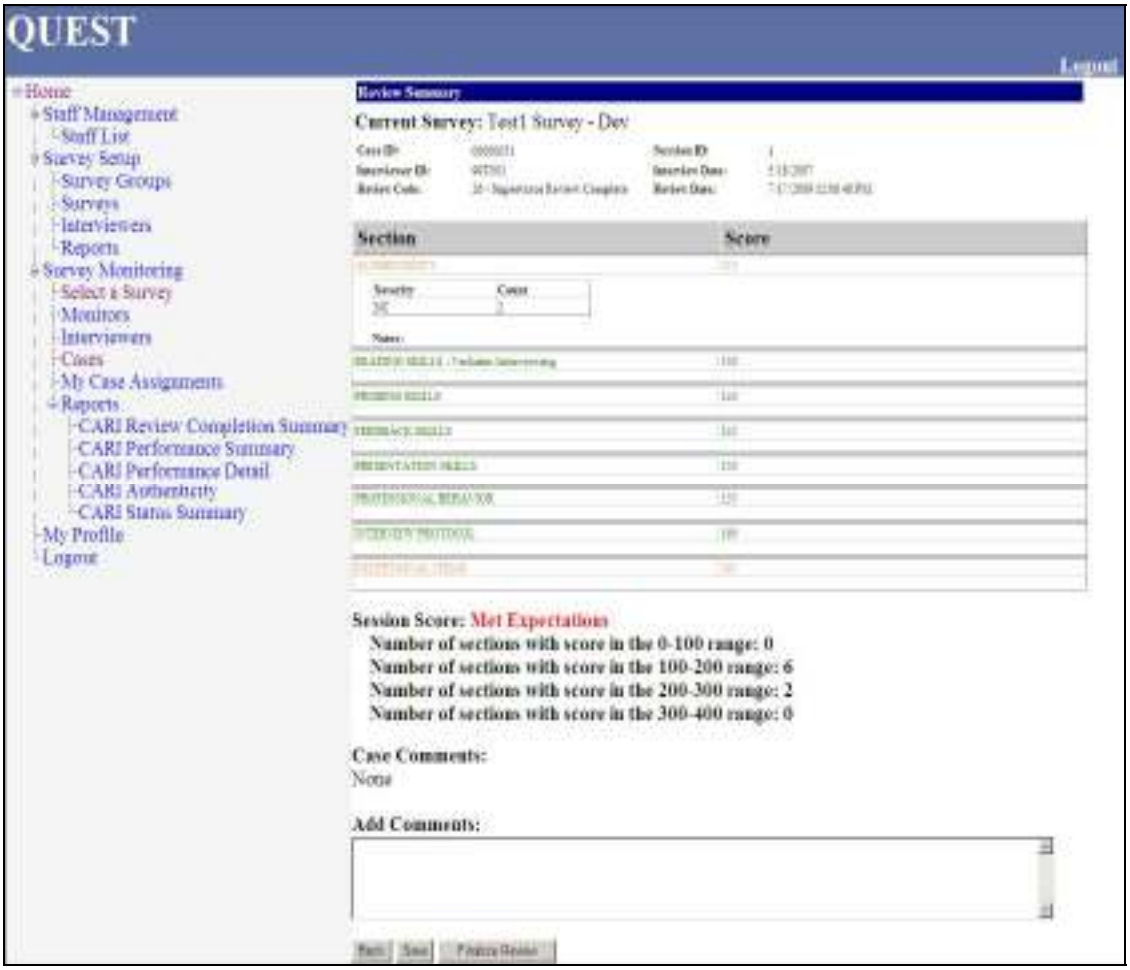
Figure 3 shows an image of the evaluation screen with the Reading Skills and Probing blocks expanded to display details. Other blocks can be expanded or collapsed at the discretion of the user, to avoid excessive scrolling. By using the “Do not consider” checkbox that appears in each criteria block, the monitor can indicate that the criteria do not apply. For example, if the audio file has recorded a series of simple yes/no questions, the Probing block may not be needed. In Figure 3, the Reading Skills block has been marked in this way.

Figure 3. QUEST Case Review Page with Mock Data



When monitors finish reviewing sound files for a particular case, QUEST displays a Review Summary Page, as shown in Figure 4, indicating the numeric score computed from the review data. In this example, several faults have been marked during the review to illustrate the workings of the page.

Figure 4. QUEST Review Summary Page with Mock Data



Tools for Managing Data Flow

The QUEST system, like other types of case management tools, depends on connectivity with other survey systems. To achieve interoperability, each survey schedules a batch job, which may run as often as hourly or as infrequently as weekly depending on the survey, writing CAPI or CATI data to the QUEST database. For CAPI projects, the audio files are delivered along with the case data and paradata. For CATI projects, audio files are delivered by the VOIP system through a script triggered by the end of the telephone call; a third action populates a table in the QUEST database with case identification information from the CATI system.

A small amount of paradata or response data, such as respondent gender, is made visible through QUEST to aid monitors with the review process. For example, if the respondent's gender is listed as male but the respondent's voice in the audio file sounds female, that may be grounds for further investigation into authenticity. Because the case's completion status is used by the sampling process, the current status code is needed in QUEST as well.

Reports

In addition to generating reports that track the progress of monitoring operations and facilitate interviewer performance feedback, QUEST has been designed with an export utility that creates an external file of the survey's performance ratings. As noted earlier, this allows projects to generate customized reports (i.e., monitoring results by respondent type) or conduct analyses that may not be needed by other surveys. Figure 5 provides a sample QUEST Summary Report summarizing the faults observed during a mock monitoring session for an interviewer.

Figure 5. QUEST Summary Report Layout with Mock Data (NOTE: This figure to be updated)

CARI Performance Summary

Survey: Test1 Survey - Dev

Report Date: 8/25/2009

Project total for each FR

FRID	Performance Code	Description	Perfomance Case Count	Total Cases	Percent
99T001	1	FI/TI and R have the same voice	1	4	25.0
99T001	2	Did not hear 2 voices	1	4	25.0
99T001	4	Problem sheet missing/inadequate	1	4	25.0
99T001	5	Callback/appt not set/inaccurate	1	4	25.0
99T001	33	Vague/inappropriate answers ignored	1	4	25.0

Confidentiality and Security

The QUEST system has been designed to protect the audio data and the confidentiality of the respondent. The system is accessible only to a restricted set of users and different levels of security are assigned based on project defined user roles and responsibilities. Users may have the role of monitor (able to see only those cases assigned to them), supervisor (able to see surveys assigned to them and appropriate reports), or administrator (able to set up new surveys and users). To prevent access to data outside the QUEST interface, audio files are stored on restricted access network shares. Audio files captured on field staff laptops are encrypted until the files rest securely behind RTI's firewall. No personally identifying information about the respondents is stored or displayed in QUEST, although confidential information could be contained in the audio recording.

Storage and Retention of Audio Files
Each survey must consider the tradeoff between having long-term access to audio recordings and the cost of storing those files. One minute of speech recorded in wave format with 16 bit monophonic quality consumes approximately one megabyte of disk space. Compression can reduce that amount by an order of magnitude (Thissen et al., 2008a), but even compressed, the amount of storage required is not trivial. Retaining the files for a short length of time, such as one week, minimizes the storage expense but at the cost of lost opportunity for review. Longer retention online allows the review process to extend over months or even years. This may be beneficial to allow retrospective reviews in the case of suspected falsification. Longer retention may also allow cross-study or longitudinal methodological research and analyses. Thus, the value of capturing and retaining files may outweigh the cost of doing so.

5. Deploying QUEST for Production Monitoring

CARI is a proven technology for capturing audio files in the field and monitoring the quality of in-person interviews, with recordings used both to authenticate interviews and support interviewer performance evaluation. As noted earlier, however, use of recorded interviews is a relatively new method for monitoring the performance of RTI's telephone interviewers. A principle design consideration for QUEST was for the system to support monitoring of telephone interviews with the features currently available through live monitoring. These include the ability to monitor respondent engagement activities such as cooperation gaining, refusal aversion attempts, and setting of appointments. Telephone quality assurance monitors, using the live audio/visual monitoring system, also rely on the observation of key strokes entered by interviewers during the interview. Although QUEST initially only captures audio recording, video capture is a planned enhancement to the system to support the evaluation of interviewer screen navigation and coding of survey responses.

Implement QUEST
RTI's call center will make a gradual transition from live monitoring to monitoring of recorded interactions in 2010, following extensive testing of the system in fall 2009. Initially, as with field surveys, only recordings from completed interviews will be evaluated in the system. Although RTI's VOIP system will capture 100% of all calls, monitoring of pre-interview contacts (i.e., screening calls and appointments) and final non-interviews (i.e., final refusals and ineligibles) will not be initiated until QUEST has been fully tested with completed interviews. During the transition period, live monitoring will continue in the call center in parallel with monitoring of recorded interviews via QUEST. The live monitoring system, however, will use the same evaluation form and interviewer performance databases so that the data can be merged. This dual-mode approach will allow for efficient yet rigorous testing of the QUEST system in our call center environment.

RTI will make a similar, gradual transition to QUEST for CARI monitoring of its field surveys. QUEST will be deployed for production monitoring on a large RTI field survey in October 2009, with additional field surveys transitioning to QUEST in early 2010.

Data collected during the transition phases will be analyzed and used to refine and/or validate the recording and monitoring parameters for both telephone and field surveys.

Collect Quantitative Data about Respondent Reactions to Interview Recordings

Field surveys have provided a wealth of data from which to gauge respondent reaction to interview recordings. In two national field studies using CARI, for example, approximately 83% of respondents in one survey of a highly sensitive topic agreed to be recorded, independent of their consent for the interview, and 93% in another less sensitive survey agreed (Wrenn-Yorker &Thissen, 2005). Other more recent RTI surveys have experienced CARI consent rates as high as 96%. Additionally, a post-interview survey of 283 respondents found that 70% of the respondents reported they had no reaction one way or another to the use of CARI, 15% reported liking the idea, while 13% disliked the idea (Herget et al., 2001).

Respondent reactions to digital recording of telephone interviews have not yet been examined in survey literature, and one concern expressed is the impact of survey introductory text that notifies respondents their call might be recorded. RTI modified standard notification verbiage for all telephone surveys in 2008, informing respondents that "this call may be monitored or recorded for quality assurance purposes." Notification is provided immediately preceding identification of the interviewer, RTI, and the purpose of the call.

With this approach, active consent to monitor or record calls is not required and doing so may depress consent rates. Passive consent (embedded in introductions) is believed to significantly reduce monitoring/recording refusal rates. With over 10,000 recorded calls to date in RTI's call center, anecdotal evidence suggests that few, if any, negative changes in respondent behavior have occurred as a result of this notification. For example, there have been no respondent objections or refusals to the recording notification. However, quantitative data gauging respondent reactions to the notification and their subsequent behavior during the interview have not been collected and literature on this subject is sparse. As digital recording becomes the standard by which monitoring is conducted, additional methodological research on this issue will be performed to better understand the impact on both telephone and field surveys.

Measure End User Resistance to Shift in Monitoring Approaches

The move from live monitoring to monitoring via recorded interactions is a fundamental shift in approach in RTI's call center. Similarly, we envision that more field surveys will use CARI for quality monitoring with the deployment of QUEST. In general, field interviewers have accepted the CARI technology. For example, in a feedback study, 82% of interviewers felt neutral or positive about the use of CARI (Herget et al., 2001). However, because the standardized monitoring approach is new, we anticipate some resistance on the part of interviewers, supervisors, quality monitors, and potentially project staff and/or clients. Collecting the necessary data to validate the system's performance, gauge reactions to QUEST, and ensure that interviewer performance feedback is timely, accurate, and fair will allow users of the system to feel comfortable with the shift in methodologies.

Increase the Utility of QUEST

The QUEST system offers a platform on which to build additional quality assurance activities. In the call center, efforts are underway to enhance the QUEST system to support quality assurance monitoring for in-bound survey calls, help desk and technical support operations, contacts and interviews with businesses and institutions, and tracing of sample members. It is important to recognize, however, that a different set of skills may be required for call center staff engaged in these tasks. For example, staff may need to exhibit a more conversational and less scripted style of interaction, a greater ability to use FAQs and other support materials to address questions from callers, or more investigative skill in identifying and selecting an appropriate informant in an institutional setting or when tracing. These skills and behaviors will be defined and the resulting evaluation criteria incorporated into the QUEST monitoring system, likely in the form of distinct item blocks tailored to these less-traditional call center functions.

Efforts are also underway to increase the number of field surveys using CARI for quality monitoring, thereby reducing the need for other monitoring methods, such as telephone verification interviews. As noted earlier, field surveys are also being encouraged to both increase and vary the audio they capture by recording consent, incentive payment, and other interactions beyond the actual survey administration. As with the call center, the QUEST evaluation criteria will be expanded, as needed, to facilitate monitoring of additional field data collection tasks. For example, one can envision the recording of sample member screenings and other door-step interactions conducted using handheld electronic devices that may be amendable to audio recording.

Finally, functionality will be added to QUEST to enable projects to evaluate the performance of individual survey items within and across instruments, yielding data to inform decisions about instrument design, the cognitive demands of specific questions, and the content of interviewer training programs.

6. Future Directions

The QUEST system replaces existing interviewer quality monitoring and reporting functionality for survey projects at RTI. The system does not significantly change the quality monitoring and control procedures but the emphasis on recorded interviewer-respondent interactions provides a richer set of data for monitoring quality and assessing individual performance. The audio recordings allow monitors and supervisors to give interviewers more useful and concrete feedback, and interviewers can hear in their own voices the behaviors noted by quality monitors. The recorded survey interactions also provide tangible evidence of survey instrument performance for our clients. These benefits represent immediate opportunities for improving interviewer performance but we also expect to achieve significant quality gains by leveraging the information collected in the QUEST system over time.

QUEST establishes a common quality assurance framework for evaluating both in-person and telephone interviewer performance. By establishing a standard set of evaluation criteria and

storing this information in a shared database, we hope to use experience gained over a large number and wide variety of projects to focus our quality improvement efforts. Collecting interviewer-level performance data over time will help us develop improved interviewer skill profiles and make better project assignments based on these profiles. Analyzing item-level performance data collected across surveys will assist our clients to develop more effective survey instruments. We also anticipate that these data will influence the techniques and approaches used to train interviewers and that trainers will frequently use the survey interaction recordings as examples.

As audio analysis software and techniques evolve, we anticipate deploying automated quality assurance checks to replace manual steps currently required. For example, RTI is working on developing an automated review of audio files to authenticate interviews. Certain elements of the interviewer-respondent exchange can be detected with an automated analysis of the voice files to verify that the interview did in fact occur. As voice recognition technology advances we also anticipate that additional behavior monitoring routines can be automated, thereby directing a smaller subset of interviewer-respondent interactions for more thorough review.

References

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

Herget, D. R., Beimer, P., Morton, J., & Sand, K. (2001). *CARI: Additional feasibility efforts of monitoring field interviewer performance.* Paper presented at the Federal Committee on Statistical Methodology Conference, Arlington, VA.

Tarnai, J. (2007). *Monitoring CATI interviews*, Presented at the 62nd Conference of the American Association for Public Opinion Research (AAPOR), Anaheim, CA.

Thissen, M. R., Sattaluri, S., Barber, L. B., & Fisher, C. (2008a). Computer audio-recorded interviewing (CARI) as a tool for monitoring field interviewers and improving field data collection. In *Proceedings of the International Methodology Symposium 2008*, Statistics Canada, Gatineau, Canada.

Thissen, M. R., Sattaluri, S., McFarlane, E. S., & Biemer, P. P. (2008b). The evolution of audio recording in field surveys. *Journal of Survey Practice*. http://surveypractice.org.

Thissen, M. R., & Wrenn-Yorker, C. M. (2005). *Computer audio recorded interviewing (CARI) technology.* Presented at the 2005 Conference on Federal Computer Assisted Survey Information Collection (FedCASIC).