QUALITY CONTROL OF CATI OPERATIONS IN STATISTICS CANADA

Walter Mudryk, M. Jane Burgess and Peter Xiao
Contact Author: Walter Mudryk, Statistics Canada, 3-D Coats Bldg., Ottawa, Ontario, K1A 0T6

KEY WORDS: Interviewer Monitoring, Statistical Process Control (SPC), Control Chart, Pareto Chart

Abstract:
This paper outlines the Quality Control methodology that has been developed for monitoring centralized CATI operations at Statistics Canada. This QC methodology is generic in nature and is based on the principles of Statistical Process Control (SPC). The main principles, concepts and design considerations underlying the methodology are presented. The paper discusses the objectives of the QC design, the quality characteristics subject to monitoring, the sample design considerations for the monitoring sample, the quality indicators that are used to measure interviewer performance, as well as, the control charts developed for tracking and evaluating interviewer performance over time. The associated Quality Control procedure relies heavily on Interviewer and Supervisor feedback which is conducted at individual levels, as well as in group settings. Special QC reports have been developed to support this feedback function that help to generate continuous quality improvement into the overall process.

The complete QC procedure along with the outputs of the 'CATI: QC Feedback System' are outlined. This includes examples of the Interviewer and Supervisor Feedback Reports which contain the control charts, Pareto charts and operational summaries that are generated for each CATI operation.

1. Introduction

During the last several years, there has been a shift toward the integration of operational processes due to the emergence of technologies with greater capabilities. This has certainly been true for survey data collection and capture processes such as Computer Assisted Telephone Interviewing (CATI), where improvements in telephone and computer technology have enabled the interviewing, editing and capture processes to come together, thereby reducing the number of steps involved in this activity. This has provided many benefits in the survey taking process, which are well known to users of CATI operations. These include the elimination of the paper questionnaire, reductions in down-line editing, lowering of respondent burden and improvements in overall survey turnaround time (i.e., cycle time). CATI is therefore considered to be both cost effective and timely as a method of data collection. The impact however, of this integration on interviewer quality is unknown, but considering some of the new complexities associated with these operations, it is likely that the potential for errors continues to exist and that in some cases, it may even be greater than before (Groves and Nicholls II, 1986).

CATI operations require interviewers to be proficient in a number of skill areas. Specifically, the interviewer must possess good basic interviewing and keyboarding skills and have a good understanding of the subject matter concepts related to the survey. The interviewer must also be capable of resolving any on-line edit failures that occur, categorize the responses, enter them accurately and be effective in maintaining a good rapport and pace with the respondent. All these things must be done well and on the fly during a live production interview.

One of the advantages of a centralized interviewing environment such as CATI, is the opportunity it provides for monitoring interviewers to determine their level of interviewing quality and enable improvements to be made to the process. CATI monitoring provides useful information for interviewer feedback, operational management, training and instrument design development (i.e., automated questionnaire), all of which contribute to significant improvements to the overall data collection process. In the past, monitoring at Statistics Canada was performed on a rather *ad hoc* basis for most head office CATI operations. Under this approach, all interviewers were monitored initially and then more sporadically unless problems arose, at which point more frequent monitoring was performed. Most applications used audio-visual monitoring while others used audio monitoring only. In the case of production monitoring, a full time staff was normally available for monitoring interviewers on some regular basis. This monitoring, even though extremely valuable for purposes of feedback and quality improvement, was not based on any statistical framework and no quantitative measures were being taken of interviewer performance. This resulted in a monitoring procedure that was rather arbitrary and inefficient from a cost perspective. Monitoring within a statistical framework, however, provides many additional advantages that enable the monitoring process to be more objective, as well as more efficient.

Statistical monitoring involves a quantitative approach to measuring quality that results in objective measures of interviewer performance. This enables the feedback to be more effective and provides the opportunity for tracking and statistically analysing this performance over time. It subsequently provides an overall framework for controlling the quality of the entire operation. Management is then better able to focus its scarce monitoring resources to areas where they are needed most, thereby making the monitoring process much more cost efficient.

This paper describes such a statistically based quality control methodology which was developed for monitoring CATI operations in Statistics Canada, using the audio-visual method of monitoring.

2. Objectives of QC Monitoring

The QC methodology was designed specifically to meet the following objectives:

(1) To provide **feedback** to CATI interviewers on their performance for their personal reinforcement and improvement.

(2) To develop a **tool for supervisors** to help them manage CATI operations more effectively.

(3) To provide an objective statistical measure of **interviewer performance** and behaviour.

(4) To indicate when and where additional **training** is required during production and to input more generally into the overall CATI training program.

(5) To ensure the maintenance of a high standard of **work ethics** in CATI operations.

(6) To provide an **assurance of quality** on the collection activities of CATI operations.

(7) To identify potential problem areas in the **instrument design.**

In general, the QC monitoring procedure is intended to **measure and improve the quality** of the entire CATI operation in terms of interviewer training, operational procedures, instrument design, survey response rates, interviewer performance and data processing.

3. Monitoring Characteristics

A first step in developing a statistical monitoring scheme involves determining which behaviours should be monitored and defining exactly how they should be measured. Accordingly, the CATI process was analysed in detail to establish the desirable and undesirable quality characteristics associated with the interviewing aspects of the process. This broadly identified all the characteristics that should be monitored and measured. These characteristics were then further analysed and classified into highly correlated error categories, in such a way that would minimize the number of categories to be monitored and thus simplify the monitoring process. A total of 11 error categories falling under six major interviewing functions were identified as follows:

Interviewing Function	Error Category
Question Delivery	Asking Wording Professionalism/Voice
Respondent Relations	Probing/Anticipation Judgement
Subject Matter	Definitions Concepts
Data Processing	Data Entry Notes
Behavioural Coding	Coding
Other	Catch All

Each major error category was then sub-divided into numerous interviewer behaviour quality characteristics. Definitions were established for each specific quality characteristic within each error category. These error categories were then placed on a 'QC Monitoring Form' so that they could be tallied during the monitoring function.

4. Nature of Check: Monitoring

The nature of check refers to the source to which the sampling unit must be compared and the actual method used to make the assessment of quality of the essential characteristics in the process. For the purpose of CATI operations, the audio-visual method of monitoring was selected as the preferred approach, following a study of several monitoring options (Mudryk, 1993).

Currently, head office operations in Statistics Canada have the audio-visual equipment to perform this type of monitoring. It is anticipated that the regional offices will be moving in this direction at some point in the near future. Therefore, the QC methodology contained in this report is structured around the **audio-visual method** of monitoring. It should be noted however, that this QC methodology can easily be adapted to support audio-only monitoring applications, with minor modifications.

The audio-visual method essentially involves a third person (i.e., a monitor) who listens, observes and assesses the interaction between the interviewer and the respondent, while working with a telephone listening device and a simultaneous duplicate image of the interviewer's computer screen. One of the main advantages of this method is that monitoring can be performed anonymously by anyone who is properly trained, thereby providing flexibility to the operation. The monitoring function is usually performed by staff who have been trained using the specially developed Self-directed Training Course for Monitors (Statistics Canada, 1995). During a given monitoring session, if an interviewer makes an error for any question or screen, this is recorded by the monitor on the 'QC Monitoring Form'. This information is then used to provide feedback, estimate quality and analyse trends and stability in interviewer performance.

5. QC Approach

Of the different methods of Statistical Quality Control available, it was felt that Statistical Process Control (SPC) would be most appropriate for controlling the CATI operations. This was primarily because the process stability assumption which is necessary for SPC to work, can generally be satisfied in CATI operations. Since interviewers undergo an initial extensive training program followed by testing to ensure that they have achieved a desired level of competence, a large degree of process stability (at the interviewer level) can be assumed at the outset. Furthermore, process control typically involves the sampling and evaluation of observations at regular intervals of time. This is highly conducive to operations such as CATI where interviewers would typically improve over time with experience and regular feedback.

The SPC procedure essentially uses **control chart theory** to evaluate process stability by comparing each sample observation against pre-determined control limits on a control chart (Shewhart, 1931). Generally, if the sample observation falls within these control limits, no action is taken. If on the other hand, the observation falls outside these limits, the process is investigated to eliminate the root cause for this large deviation. This would occur for example, when a critical error (or its equivalent) is discovered by the monitor, which must then be immediately fed back to the interviewer for corrective action. Once this root cause has been addressed, process control sampling resumes for that interviewer.

It should be noted that under this approach each interviewer is considered to be an **individual process** in their own right since firstly, the interviewer is considered to be a major source of error variability and secondly, feedback must be provided at this level. Therefore, a separate control chart is maintained for each interviewer.

6. QC Design Considerations

The following are the statistical design considerations that form the basis of the generic QC methodology which is being used for monitoring our CATI operations:

6.1 Sampling Unit

The *sampling unit* in quality control is usually the unit of product or part of the process over which we want to establish some form of control. For CATI operations, it was decided that this should be a function of an interview, since it is the **interviewing process** that we desire to control. Accordingly, the following units were considered to be good candidate sampling units:

- Complete Interview
- Partial Interview
    - Time Interval
    - Question(s)
    - Screen Display Information

A **complete interview** allows the monitor to experience the full interaction between the interviewer and the respondent. This permits the monitor to gather accurate information about the situations that are not handled well by the interviewer. However, if a complete interview is used as the sampling unit, the monitoring session could turn out to be quite long for some surveys and hence a smaller number of interviews would have to be monitored per interviewer. These longer monitoring sessions could also cause fatigue to the monitor because the level of concentration required by the monitor is quite high and may not be sustainable over extended periods of time (e.g., say for an hour or more). This could ultimately reduce the efficiency of monitoring. Another major disadvantage with this choice of unit is the difficulty of being able to get a 'complete' interview as a unit of observation, since the monitor will rarely be able to select an interviewer right at the beginning of an interview (i.e., most interviews are already in progress when selected).

The other candidate sampling units come under the broad heading of **partial interviews**. These units represent shorter monitoring sessions that have the distinct advantage of allowing a sufficiently large number of quality characteristics to be monitored within a reasonable period of time. This tends to increase monitoring efficiency. Furthermore, because it allows one to begin at any point in the interview, it leads to significantly simpler sample selection procedures.

Considering the first candidate unit under the partial interview category, namely time interval (e.g., 15 minutes of monitoring), the number of observations monitored for different interviewers can be highly variable and this would require more complicated procedures for assessing and reporting quality. Conversely, if the sampling unit is a question or screen display, it would allow for the selection of a fixed number of observations (i.e., **fixed areas of opportunity**) for each interviewer. This would subsequently lead to simpler sample selection procedures and alleviate the complexities associated with assessing quality from fluctuating sample sizes and subsequently estimating interviewer performance. Furthermore, the '**screen display**' permits the measurement of additional interviewing characteristics such as 'introductions' and 'closings', with a fixed sample size. Therefore, a complete computer 'screen display' of interviewing activity was considered to be the most appropriate choice of *sampling unit* for monitoring our CATI applications.

In the context of CATI operations, a '**screen**' is defined as any piece of interviewer activity involving a respondent, which occupies the full screen on a computer terminal. The screen can contain for example, a question (with or without sub-parts), an introduction, closing, call back arrangements, etc.

6.2 Sample Selection

The purpose of sample selection in CATI monitoring is to obtain a good representation of an interviewer's quality (i.e., their interviewing skills) at specific points in time, so that these observations can then be evaluated for process stability over time. This is best achieved by selecting periodic **subgroup samples** from the process (Wheeler and Chambers, 1986). A random **cluster** sample of screens (i.e., a rational subgroup sample) is selected for each interviewer at regular intervals of time and plotted on a control chart. This then permits an evaluation of process stability to be made on the interviewer's performance by determining how consistently these subgroup samples behave over time. For CATI operations, this is achieved as follows:

- An interviewer is selected from the group of **active** interviewers and a specified number of **consecutive sample screens** (i.e., starting from the screen being worked on) is then monitored for that interviewer.
- After having completed the monitoring for that interviewer, another interviewer is selected, once again based on him or her being active on the telephone at that particular time.
- Additional samples are then selected for each interviewer throughout that week at approximately even intervals of time, as specified by the sampling plan requirements for that interviewer.
- Selection of samples is controlled by completing a 'Sample Control Form' each time a sample is selected and comparing this against the sampling requirements for each interviewer for that week.
- If the required number of screens cannot be achieved in a given monitoring session, a number would be noted on the form which would tell the monitor that additional screens for that interviewer must be selected on the next available occasion. The combination of these two sessions would then form the sample observation for that interviewer.
- This procedure is repeated across all interviewers until the required number of samples is selected for each interviewer for that week.

This procedure approximates **systematic random cluster sampling** of screens within an interviewer, since the combination of interviewer and monitor availability, at any point in time, will occur on a random basis and interviewers are then sampled by selecting fixed clusters of output (i.e., screens) at approximately even intervals of time within a given week.

Another important criterion taken into account is that all interviewers are not sampled at the same rate over time. This is because some interviewers are experienced and do not require the same amount of monitoring as others. Accordingly, interviewers are **stratified** into four groups based on their previous weeks performance, as follows: (A) experienced or excellent interviewers (B) very good interviewers (C) acceptable interviewers and (R) unacceptable performers. After a suitable training and testing period, all **qualified interviewers** are placed into group C. Over time, as interviewers improve their skills and abilities, they are moved from group C to B and likewise, from group B to A. Those that deteriorate to an unacceptable level however, are moved to group R. The rules for movement from one group to another are described in Section 7 below. These four groups are then sampled at different rates at approximately fixed intervals of time. Group A interviewers are monitored at two subgroup samples per week, group B interviewers at four subgroup samples per week, group C at six subgroup samples per week and group R at a minimum of eight subgroup samples per week.

6.3 Sample Size

As previously stated, the purpose of sample selection is to obtain an indication of each interviewer's performance at specific points in time, based on a series of relatively small rational subgroup samples. These sample observations are subsequently plotted on control charts and evaluated for their stability to determine if the interviewers' variation within each subgroup sample is consistent with the variation between the subgroup samples (e.g., variation within period vs. variation between periods of observation). It is the accumulation of these subgroup samples, plotted over time on a control chart, that provides the basis for this evaluation.

Several factors have highly influenced what this subgroup sample size should be. These include a) the number of quality characteristics to be measured, b) the type of quality measure to be used and c) the limited amount of monitoring resources available. For CATI operations, many different characteristics are monitored for each screen (i.e., currently classified into 11 error categories), giving rise to many opportunities for error, not only within screen, but also across screens or subgroups of the interview. Since the quality measure being used is a count of **all** defects encountered within this sample (i.e., multiple defects within screen or repeating defects across screens are all counted), a moderate subgroup sample size is sufficient to obtain a meaningful measure of quality of the interviewer's performance. Furthermore, since monitoring resources are scarce (i.e., about one monitor available per staff of 15 - 20 interviewers), it was felt that a fixed subgroup sample of 20 consecutive screens, would be sufficient to act as a meaningful **benchmark measure** of an interviewer's quality while at the same time, enabling effective process decisions to be made. In practice, this subgroup size has worked quite well for our applications.

6.4 Quality Measure

In every quality control situation, there must be a suitable quantitative measure of quality that will account for certain important features of the product or process being measured. In the case of CATI operations, this quality measure relates to the interviewing ability of an interviewer. This measure must therefore be capable of expressing the degree of acceptability of an interviewer's performance in terms of their interviewing.

A quality measure can be defined in many ways but in the final analysis, it must be simple, meaningful to the user and effective in measuring and expressing the desired quality characteristics of the process. Furthermore, the cumulative effect of all errors is extremely important in determining the overall quality of the interviewing process. Since the sample size must be relatively small, a quality measure using all available information should be used. Therefore the quality measure known as "**defects per fixed area of opportunity**" was selected (Wheeler and Chambers, 1986). In the CATI context, this is the total count of all defects observed per fixed number of screens (i.e., in our case, 20 screens).

Now suppose k monitoring sessions are observed for interviewer j in week i. Under this approach the weekly quality measure for interviewer j is given by:

$$c_{ij} = \sum_{l=1}^{k} \frac{c_{ijl}}{k}$$

where c_{ijl} is the total number of defects (i.e., unweighted) in the l^{th} monitoring session of 20 screens. The quality measure, c_{ij} is the average count of the number of errors per 'sample' for interviewer j in week i.

The average weekly quality measure (based on unweighted defects) for the entire operation, over all interviewers is then given by:

$$c_i = \sum_j w_{ij} c_{ij}$$

where w_{ij} is the appropriate volume weight of each interviewer's contribution to the total operation during week i. This is given by $w_{ij} = N_{ij}/N_i$ where N_{ij} is the total number of screens processed by the j^{th} interviewer in week i and N_i is the total number of screens processed by all interviewers during week i.

6.5 Seriousness Classification of Errors

In the development of the Quality Measure, it was recognized that some errors are more serious than others. Therefore, a seriousness classification system for recording different types of errors was established to help classify them. This system allows one to assign a relative weight to each different type of error condition and thereby categorize it according to its relative seriousness. For CATI operations, a demerit rating system (Juran, 1988) was selected as follows:

Critical Error = 4 demerits
Major Error = 2 demerits
Minor Error = 1 demerit

The use of demerits alleviates the problem of having fractional error weights and is simple to implement operationally. The application of this demerit rating system is desirable for purposes of interviewer quality estimation and sampling plan qualifications. It allows an evaluation of interviewer performance on a **critical error equivalent basis**, whereby an accumulation of enough minor and/or major errors would be equivalent to one or more critical errors.

For purposes of quality estimation, a demerit weight is assigned to each c_{ijl} component in such a manner that the interviewers' quality measure (i.e., total weighted error count per sample) becomes the total sum of these demerit errors per 20 screens. Therefore, the **weighted demerit rate per sample** can be represented by c^*_{ijl} instead of c_{ijl}. Specifically, $c^*_{ijl} = \sum_h W_h \, c_{ijlh}$ where W_h is the demerit weight for the h^{th} class of error (here h = 1 to 3 represents the minor, major and critical classes of error, respectively). The corresponding **weighted quality measures** for the interviewer and the overall operation, expressed in demerits per sample, are then given by:

$$c^*_{ij} = \sum_{l=1}^{k} \frac{c^*_{ijl}}{k} \quad and \quad c^*_{i} = \sum_{j} w_{ij} \, c^*_{ij}$$

6.6 Control Charts

As previously stated, since the critical errors (or their demerit equivalent) are automatically being flagged and responded to under the current feedback procedures, and since it is the demerit counts that determine which observations are out-of-control, the **d-chart** was selected as the most appropriate model to use in our case (Pyzdek, 1993). This type of control chart is applicable when the quality measure is an attributive count of demerits with a constant sample size. The d-chart is sample size independent, relatively simple to implement and easy to understand. In this case, since we are dealing with three classes of error with different demerit weights, the mixed Poisson distribution would apply. This is primarily because we are dealing with discrete error counts coming from a fixed sample size, whose general frequency of occurrence is expected to be quite low.

In adapting the d-chart to the CATI operations, several decisions had to be made. Typically, the *centre line* of a control chart is taken as the process quality (e.g., process average, number of defects or demerits per sample, etc.) and the *control limits* are calculated at three standard deviations away from this centre line. In our case, however, since observations containing any critical errors (or their equivalent of 4 demerits or more) are flagged automatically and considered as points out-of-control, it was decided to apply ***user specified control limits*** due to this **operational constraint**, as opposed to computing them. Since a critical error has a value of 4 demerits, this was logically selected as the user specified value for the Upper Control Limit (UCL). Furthermore, since extremely low error rates were considered to be good indicators of process quality and favourable observations which are in-control, a Lower Control Limit was considered to be unnecessary and therefore this value was set to zero (i.e., LCL = 0). Now since the control limits were not being calculated, the centre line was set at the weekly group average demerit rate across all interviewers. It was felt that this would represent a good benchmark of process quality for the interviewers to compare themselves against. Constructing the control chart in this manner provides several advantages. It alleviates the problem of having many different centre lines and control limits for the interviewers and also reduces the amount of computations that would otherwise be required. The individual interviewer sample observations (i.e., c^*_{ijl}) are then plotted on this **standardized control chart** and evaluated against the pre-established control limits for the operation. In this manner, all interviewers are essentially judged against the **same standard**, as opposed to their own personal control limits, which could vary substantially across interviewers. This also alleviates any potential confusion that could exist among interviewers relating to the administration of variable control limits.

These modified control charts are essentially more sensitive to process requirements (i.e., user specifications) than to process stability in the sense that, regardless of the process variation, if a sample observation falls above the user specified control limit, it is judged to be unacceptable (and out-of-control) and an investigation is immediately launched to determine the root cause so that appropriate action can be taken. Therefore, if the process is not capable of meeting this minimum acceptable client specification, these charts become more responsive in identifying this than the normal SPC charts would be.

The information for the control chart is obtained from the QC Monitoring Forms which are input into the CATI: QC Feedback System. This system automatically generates the required control charts for each interviewer which then become an essential component of their feedback package (shown in the Appendix). The system also generates additional information for these feedback reports (e.g., Pareto charts and operational summaries) which are discussed in Section 7 below and are also shown in the Appendix.

6.7 Pareto Analysis Charts

During the monitoring process, as errors are discovered, they are recorded on the QC Monitoring Form by error category (e.g., question asking, data entry, etc.). The weighted error frequencies are then accumulated for each category for a specific time period (e.g., one week) both within and across interviewer observations and plotted on a frequency distribution chart. This chart, known as a **Pareto Chart**, is used to show where the majority of errors have occurred within the operation. Performing this analysis at the interviewer level is especially useful for the purpose of inspiring self-improvements.

Performing the analysis across all interviewers for the entire operation is useful in identifying more general problems associated with the process (e.g., operational, procedural, systems, environmental, instrument design, etc.) which is the focus of subsequent group feedback sessions. An example of these Pareto charts are shown in the Appendix.

7. Control Chart Action Plan
This refers to the steps that are taken when the control chart data for any interviewer identifies any unusual patterns, trends or undesirable points. The following signals were determined to be worthy of some form of decisive follow-up action:

- One or more points fall on or outside the user specified upper control limit, UCL
  - If one point falls outside the UCL, immediate feedback is given to the interviewer; monitor additional samples for that interviewer as soon as possible. If two or more points fall outside the UCL within a week, inform supervisor, give feedback and remove interviewer from production for re-training.
- Two or more points fall close to the user specified upper control limit, UCL
  - Provide precautionary feedback of the situation to the interviewer.
- A run up or run down of 5 or more consecutive points
  - Re-evaluate sampling plan qualification for interviewer. If the run is down, assign a better (i.e., lower frequency) sampling plan. If the run is up, assign a sampling plan with a higher sampling frequency. See qualification rules below.
- Cyclical or other non random patterns
  - Re-evaluate sampling plan qualification for interviewer as per qualification rules below.

Qualification Rules:
The following qualification rules for sampling interviewers at different rates, are based upon pre set *Qualification Limits* (QL). These limits are essentially standards of acceptability for different levels of processing that result in the use of different sampling plans. Generally speaking, if these limits are satisfied, interviewers are qualified onto the appropriate sampling plan. Currently, the pre-set qualification limits are: QL= 4 demerits for Plan C, QL=3 demerits for Plan B and QL=2 demerits for Plan A. Using these limits, the following qualification rules were developed:

a) After an initial training and testing period, all interviewers are automatically qualified onto Plan C. Here they are sampled a minimum of 6 times during the first week (i.e., approximately once per day and more if necessary). If a run of their last 5 samples each contain less than or equal to 3 demerits (i.e., QL=3), the interviewer is qualified onto Plan B.

b) On Plan B, interviewers are sampled 4 times per week. If a run of their last 5 samples each contains less than or equal to 2 demerits (i.e., QL=2), the interviewer is qualified onto Plan A.

c) On Plan A, interviewers are sampled 2 times per week. If a single point is greater than or equal to 4 demerits, the interviewer is moved back to Plan B. A minimum of 5 additional samples is then required to re-qualify onto Plan A.

d) If 2 of the last 3 samples for a Plan C interviewer, contains 4 or more demerits (i.e., QL=4), the interviewer is placed onto Plan R. On this plan, the interviewer is re-trained and additional samples are then selected until 2 consecutive samples contain less than 4 demerits, at which point, the interviewer is once again re-qualified onto Plan C.

e) If an interviewer on Plan B has a single point greater than or equal to 5 demerits, the interviewer is moved back to Plan C. Once again, a minimum of 5 additional samples is required to re-qualify onto Plan B. It should also be noted that it is possible for the interviewers to move further back onto Plan R as per rule d) above, when applicable.

For similar or repetitive short run applications being processed by the same interviewers, the sampling plan qualifications are generally carried forward to the next occasion. This allows for some continuity in interviewer qualifications for applications that are similar to one another.

8. QC Feedback
Currently, the head office operations in Statistics Canada provide regular feedback (i.e., both positive and negative) to all interviewers individually, as well as in a group setting. An essential part of the QC strategy is to provide timely feedback of results to all interviewers monitored, so that they can use this information to continuously improve their performance.

Individual feedback to interviewers is accomplished in two formats, namely, immediate feedback and weekly feedback. *Immediate feedback* is provided to interviewers whenever a critical error or its equivalent is observed. This ensures that any out-of-control point (i.e., out-of-specification) on the control charts is responded to immediately. This is usually done by the monitor as soon as the interview is finished, to ensure that important errors are prevented from recurring on subsequent occasions. In this case, the monitor provides detailed information and comments for each critical error (or its equivalent) that was recorded on the QC Monitoring Form.

Generally, *Weekly feedback* takes place at the beginning of the following week. To accomplish this, individual feedback reports are generated by the CATI: QC Feedback System which includes a control chart, a Pareto analysis chart and an operational summary for each interviewer. This information is distributed to each interviewer to encourage further self-improvements in their performance. The monitor also ensures that each interviewer receives some personal oral feedback (i.e., either positive or negative) at least once per week.

In addition, the CATI: QC Feedback System generates reports for the supervisor of the operation. These reports include a weekly comparison chart of all interviewer performance, a Group Pareto Chart of errors, as well as a Group Analysis Report which shows the distribution of interviewer sampling plan qualifications and provide estimates of quality for the last week and overall weeks of the operation. These reports can be used by the supervisor to help them determine further training requirements and manage their CATI operation better. This normally includes incorporating this information into **Group feedback** sessions which are conducted periodically by the supervisor. These sessions use the group information generated by this system to help focus on further improvements to the process which are not necessarily the fault of the interviewer. Problems related to environment, operations, specific procedures, systems, instrument wording, etc. are usually identified at this level and further improvements are then made to the overall process.

This package of feedback information is also used by other stakeholders in the CATI operation, namely, managers, designers and methodologists. Managers and designers receive an assurance of quality in the CATI interviewing process and are better able to analyse the progress of the operation in terms of outgoing quality and costs. Methodologists receive valuable information to help them evaluate the effectiveness of the QC methodology and make changes as required.

9. Summary of QC Procedure

The following is a brief overview of the QC Procedure that is used to implement the CATI: QC methodology:

(1) An interviewer is selected from the population of active interviewers for monitoring. The monitor uses the Sample Control Form to select the required number of monitoring samples per interviewer per week.

(2) For each monitoring session, the monitor records the number of screens observed in the sample on the Sample Control Form. If less than 20 screens (i.e., our current sample size) are available for the selected interview, the monitor selects additional screens from the next available interview for that interviewer, to complete this sample observation.

(3) The monitor observes 20 consecutive screens from the selected interview(s) and records the errors (with detailed comments, such as question number and nature of error) on the CATI: QC Monitoring Form. Positive feedback relating to the interview is also recorded on this form.

(4) The monitor provides immediate feedback to the interviewer and/or supervisor of any critical errors (or their equivalent) that were observed during the sample monitoring session.

(5) The monitor proceeds to select additional samples for interviewers as specified by the Sample Control Form for that week. Interviewers are always selected as needed from the population of active interviewers, at that point in time. This is repeated until all sampling requirements on the Sample Control Form are satisfied.

(6) The QC Monitoring Forms are then compiled and processed weekly by the CATI: QC Feedback System which generates the appropriate control charts, Pareto analysis and various operational summaries required for feedback to interviewers, supervisors and managers of the operation (see Appendix for all output reports).

(7) Feedback reports are distributed on a weekly basis and new sampling plan qualifications are made for each interviewer for the next cycle. This cycle is repeated for each week of production.

10. Concluding Remarks

To date, we have applied this QC methodology to approximately 35 head office survey applications with very positive results. This approach allows us to identify interviewer problems early in the CATI process and take corrective action, as necessary. In addition, the QC data is used in group feedback settings to identify additional process problems that may be affecting the operation. The combined efforts of dealing with interviewer and other related process problems early in the operation, enables us to incorporate continuous quality improvement into each CATI operation. The response from the client community has been extremely positive and subject matter specialists are pleased to know that the quality of their CATI operations is being monitored, quantified and improved over time.

We are now in the process of automating the CATI software to enable the monitors to record the QC information on-line instead of manually on the QC Monitoring Form. This will facilitate the monitoring process considerably. This information will then be automatically linked to the CATI: QC Feedback System to produce the desired output reports.

11. Acknowledgements

The authors wish to thank Doris Morrow and Bob Bougie for their valuable support in developing and implementing this CATI QC methodology into the head office operations of Statistics Canada.

The authors also wish to thank Jean-François Gosselin and Bob Bougie for their valuable comments in reviewing this paper.

12. References

Groves, R.M. and Nicholls II, W.L. (1986), "The Status of Computer-Assisted Telephone Interviewing: Part II - Data Quality Issues," *Journal of Official Statistics*, Vol. 2, No. 2, pp. 117-134.

Juran, J. M. (1988), *Juran's Quality Control Handbook*, 4th Ed., McGraw-Hill, New York, NY, pp. 18.38 - 45.

Mudryk, W. (1993), "Monitoring Options for Centralized Interviewing Operations," Internal Report, BSMD, Statistics Canada.

Pyzdek, T. (1993), "Process Control for Short and Small Runs," *Quality Progress*, ASQC, April, pp. 51-60.

Shewhart, W.A. (1931), *Economic Control of Quality of Manufacturing Processes*, ASQC, Milwaukee, WI.

Statistics Canada, (1995), "A Self-directed Training Course for Monitors of CATI Operations", Operations Research & Development Division, Ottawa, Canada.

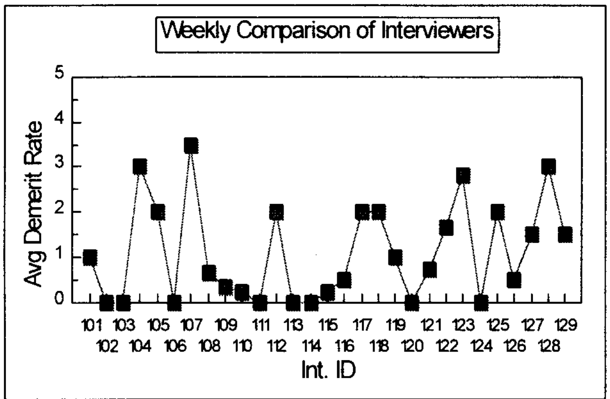
Wheeler, D.J. and Chambers, D.S. (1986), *Understanding Statistical Process Control*, SPC Press, Knoxville, TN, pp. 63, 242.

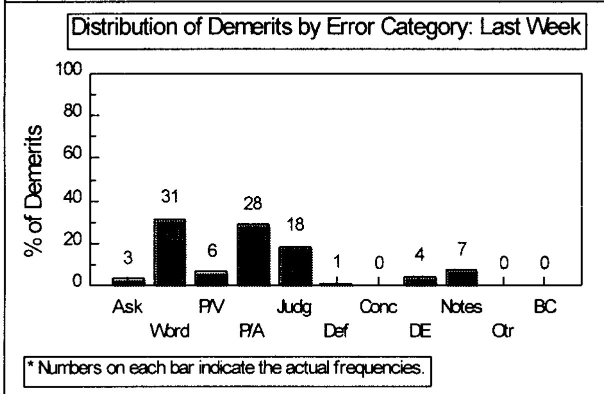
Appendix

Please refer to the next page.

Appendix

CATI: Weekly Supervisor QC Reports
Survey Name: _________





Analysis Report

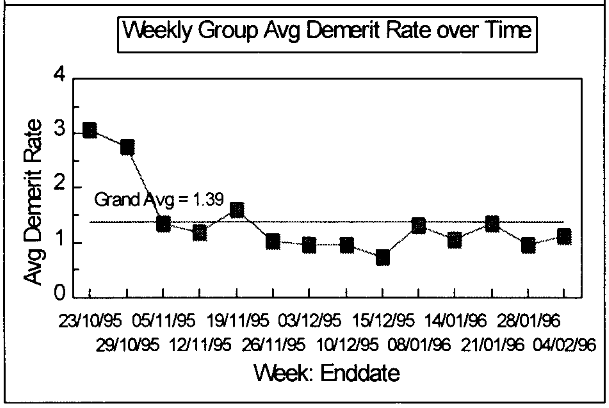
Weekly Cycle: 29/01/96 - 04/02/96

Group Rates (Per 20 Screens)

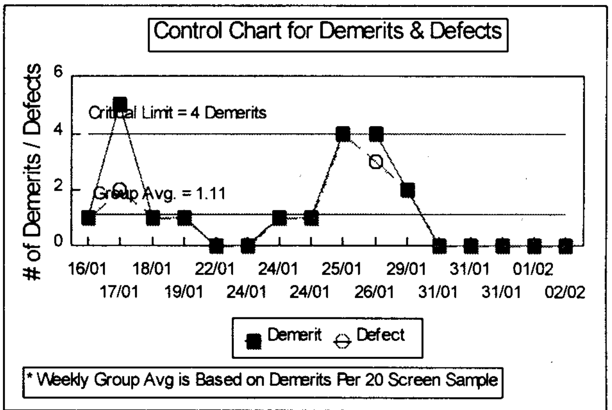
Demerit Rates:	Last Week	1.11	All Weeks:	1.39
Defect Rates:	Last Week	0.99	All Weeks:	1.18

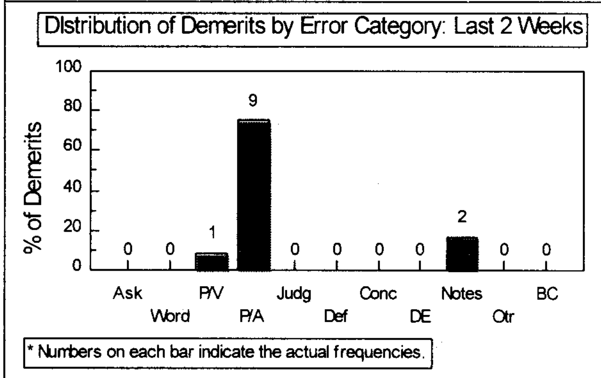
Interviewer Qualifications

Plan A:	18 Interviewers
Plan B:	7 Interviewers
Plan C:	4 Interviewers
Plan R:	0 Interviewer



CATI: Weekly Interviewer QC Reports
Survey Name: _________ Int ID: ____





Analysis Report

Weekly Cycle:	29/01/96 - 04/02/96	New Plan Qualification:	B
		Last Week Qualification:	C

AVG ERROR RATES: Last Week (Per 20 Screens)

Interviewer Defect Rate:	0.33	Interviewer Demerit Rate:	0.33
Group Defect Rate:	0.99	Group Demerit Rate:	1.11

Interviewer Demerit Rate: All Weeks:	1.18
Group Demerit Rate: All Weeks:	1.39

COMMENT: