Continuous Quality Improvement for Survey Operations:General Principles and Applications

Paul Biemer and Rachel Caspar, Research Triangle Institute
P.O. Box 12194, Research Triangle Park, NC 27709

Keywords: Continuous Quality Improvement, TQM, Industry and Occupation Coding

1. INTRODUCTION

A typical sample survey consists of a number of separate, but interrelated operations that may either change the form or modify the content of the data. These operations include data collection (or interviewing), data transmission and receipt, data editing and cleaning, response encoding, and data entry. Depending on the scale of the survey, these operations may involve only a few operators or, as in the case of a census, hundreds. It is also quite common that the operators are inexperienced, lowly-paid, and minimally trained workers. The operations in which these operators are engaged may be quite complex and error prone. They may consist of repetitive and monotonous activities or, as in the case of interviewing, may require complex thinking and quick judgements on the part of the operator. Thus, almost all survey organizations employ some type of quality control for survey operations in order to ensure that the final results are of acceptable quality.

In most survey organizations, quality control is usually based upon some form of acceptance sampling method. Acceptance sampling methods (also known as inspection methods) involve selecting a sample of items from a work unit, inspecting the sample to determine the number of items which are "in error" or which deviate from specified procedures, and then either rejecting the work unit if the number of errors (or defects) exceeds some threshold value or otherwise accepting the work unit. For most operations, rejected units are usually reworked to remove the errors.

There are at least three potential objectives of inspection methods. One is to ensure that the error in the output of an operation does not exceed some specified level. This is possible through the use of probability methods for determining the number of units to sample, the number of items within each unit to inspect, and the threshold value for rejecting a unit.[1] For some implementations of acceptance sampling, this may be the sole purpose of the activity. However, for other implementations, a second objective may be to improve the skills of the operators. This is usually carried-out by providing "feedback" to the operator responsible for a rejected work unit; that is, the operator is given information on the number and types of errors found in a rejected unit. In this way, it is hoped that the operator can take whatever steps are necessary to avoid these types of errors in the future work. A third objective may be simply to "keep the operators on their toes." That is, by providing the "threat" of inspection, the operators are dissuaded from intentionally deviating from procedures or taking "short-cuts" which may damage data quality in order to meet production goals.

This use of acceptance sampling may be criticized on several grounds. First, inspection adds significantly to the costs of a product. Inspection is necessary only because the operation is prone to error. If the process could be redesigned so that the error rate were extremely small, inspection would be unnecessary and the cost of inspection would be saved. Second, acceptance sampling can ensure an acceptable level of quality only if the inspection process is nearly perfect. However, this is seldom the case as the inspection error rate may be substantial. Thus, the error in the final product may be unacceptable even with a high rate of inspection. Further, to achieve very small levels of error - say, 1% or less - 100% inspection is required. However, this level of inspection may be unaffordable. In addition, the inspection process may miss a large proportion of the errors in the operation since the inspection operation is as large as the operation being inspected. Third, the recently quality control literature (see, for example, Ishikawa, 1990) provides strong evidence that when inspection methods are used, operators take less responsibility for the quality of their work since quality is perceived as being the job of the inspectors. Therefore, the operators lack the motivation to improve quality. Fourth, when feedback to the operators is based upon rejected units, the implication is that they are solely responsible for the errors in their work unit. However, the quality control literature suggests that for most operations, the operator may only be responsible for 20-30% of the errors. Thus, placing full responsibility for the errors in a process on the operators risks demoralizing them. This demoralizing effect is exacerbated by an imperfect inspection process which both fails to identify true errors and erroneously classifies correct items as errors. Finally, the feedback from inspection to the operators is both time consuming and, in many cases, ineffective. Part of the reason for this is the lack of information on the "root causes" of the errors. The operator may be told that he/she is responsible for various errors, but is given little if any useful information on how to eliminate them in the future. Indeed, as we have stated, many errors may be beyond the control of the operators.

These limitations of the traditional inspection methods of quality control motivated our investigation of alternative methodologies for ensuring high quality in survey operations. A key distinction between continuous quality improvement (CQI) methods and traditional quality control methods is that the former aims, not at simply achieving a specified "average outgoing quality limit" (AOQL), but at achieving the smallest error rate possible by continually improving the quality of the product for the entire duration of the operation. In this work, we drew heavily from the literature of total quality management (TQM) to identify management practices and approaches that encouraged continual improvement in the way operators performed or are able to perform their jobs. Unfortunately, the literature on applying TQM methods to survey operations is essentially non-existent so that this paper represents one of the first, albeit rudimentary, documented applications of the principles of TQM to a survey operation.

In the next section, we provide a framework for implementing CQI strategies to a wide variety of survey operations. Also discussed in this section are the fundamental concepts and tools that define our specific CQI approach. In Section 3, we describe an experiment in which our approach was implemented in an industry and occupation coding operation. In presenting the results, we compare the error rates and costs of the CQI approach with that of the traditional inspection method. Finally, in Section 4, we summarize the lessons learned in our study and discuss a number of issues related to the implementation of the CQI methodology to other operations.

2. THE BASIC CQI PRINCIPLES

2.1 A Conceptual Framework for Survey Operations

In this section, we present a general strategy for quality improvement which is applicable to a variety of survey operations. In describing this strategy, it is useful to provide a conceptual framework consisting of the fundamental components of the typical survey operation. Using this "model" of the survey operation, we will describe in general terms the objective of CQI and how these objectives can be realized.

Most survey operations consist of three major components or stages: the input (or stimulus), the action (or task), and the output (or result). The input stage, which may be the output of some previous operation, may consist of data, forms, or other information requiring some action by an operator. These input items may be assigned to an operator for processing in work units of some homogeneous size. In the action stage, the operator performs the tasks associated with the operation on the input items. The results of these actions constitute the output for the operation. Associated with each of the three components of a survey operation are the actual and the preferred inputs, actions, and outputs. As an example, the actual inputs for interviewing are the questions and procedures as they are currently defined. The preferred inputs are those questions and procedures that encourage preferred actions from the operators. Likewise, the actual actions for interviewing are the actions taken by an interviewer during an interview whereas the preferred actions are those actions which would have elicited the best response (or the preferred output). Finally, observations or interview results constitute the actual output while the preferred output are results which are completely accurate and free of nonsampling error. In brief, the actual component is what exists in the current operation and the preferred component is the ideal input, action, or output.

We shall assume that the actual survey operation component can be observed and that the preferred survey operation component can be uniquely and unambiguously defined so that actual and preferred components can be compared. The difference between "actual" and "preferred" performance for a particular item will be referred to as a nonconformity. Thus, the ultimate goal of CQI is to change the actual performance of an operation to agree perfectly with the preferred performance so that the number of nonconformities in the operation is reduced over time to zero. Progress toward this goal is achieved if, at each implementation of the operation, the number of nonconformities is reduced from the previous implementation. Thus, the objective of CQI is to continuously and incrementally move the current operation toward the preferred operation thus reducing the number of nonconformities in the operation to zero. Note that, unlike inspection methods which tend to focus only on the operator, CQI addresses all three components of the operation.

2.2 A General Strategy for CQI

Using this conceptual framework for a survey operation, in this section we propose a general strategy for implementing CQI. Our approach may be viewed as an integration of three fundamental principles of TQM. First, a critical ingredient in our approach is the use of teams to identify problems, to determine their solutions, and to implement corrective measures. Secondly, the actual components of a survey operation are evaluated quantitatively using quality indicators which are functions of the number of nonconformities in the operation. Finally, the highest priority is given to identifying and addressing the root causes of the nonconformities without regard to where they are in the system or the organization.

Our CQI plan is a four step approach and an adaption of Deming's (1986) Plan-Do-Check-Act (PDCA) cycle. However, our approach is especially adapted for survey operations and is more specific regarding the activities to be performed under each step of the cycle, particularly the "Planning" and "Checking" steps. The four steps are as follows:

Step 1. Perform the operation and observe the nonconformities.
Step 2. Classify the nonconformities as to their type and perform a Pareto analysis.
Step 3. Meet in teams to identify the root causes of the most important types of nonconformities.
Step 4. Implement the corrective measures and return to Step 1.

Step 1. Perform the operation and observe the nonconformities.

"Observing the nonconformities" implies that there is a comparison of the actual performance of the operation and the preferred performance for all three components of the operation. As an example, for telephone interviewing, the comparison may be made by a call monitor who is proficient in survey procedures and who, while listening to the interview, determines whether the observed behavior agrees with the preferred behavior. For editing, data entry, and coding, this step may entail reworking a sample of the items by an expert or by using some other process which guarantees preferred output. Inspection methods such as independently reworking a sample of items, comparing the outputs, and then adjudicating the differences to obtain a final adjudicated output may be used to produce the preferred output.

Step 2. Classify the nonconformities as to their type and perform a Pareto analysis.

Step 1 may identify many different types of nonconformities in the operation - too many to address simultaneously. This step sorts the nonconformities by type and performs a Pareto analysis, i.e., orders the type by the frequency of their occurrence.[2] This analysis will allow us to focus on a few, more important types of nonconformities in Step 3. As an example, for industry and occupation coding, the type classes for the nonconformities may correspond to the final adjudicated occupation or industry code. In this way, those occupations or industries which are particularly prone to error may be singled out for special consideration.

For interviewing, the system of monitoring proposed by Couper et al. (1991) can be used to classify the types of nonconformities observed during interviewing. These type classes correspond to nonconformities in the delivery of the question (wording changes or skipped questions), probing for an adequate response (probe neutrality, completeness, or failure to probe), interviewer feedback to the respondent (neutrality or appropriateness), respondent behaviors (requests for clarification or to repeat the question) and so on. With this system, a sample of questions is monitored and the type of nonconformity observed is coded for each question and post-question interaction. Thus the type classes for the Pareto analysis may be based upon these interviewer or respondent behavior codes for all questions on the questionnaire combined, for particular sections of the questionnaire, or for individual questions.

Likewise, type classes may be defined for coding, data entry, or data editing. This may require constructing a list of the various types of nonconformities observed in the operation and developing a classification system on the basis of the most frequently observed errors. The nonconformities may be further stratified by input, action, or output. For example, nonconformities in the input affect the appearance of the data as they are presented to the editors, keyers, coders, etc. and should be reported to the previous operation. Nonconformities observed in the output may be reported by the subsequent operations in the sequence of survey operations, but may also be observed in the current operation.

Step 3. Meet in teams to identify the root causes of the most important types of nonconformities.

A key feature of our approach is the use of teams to fully investigate the nonconformities until their root causes are well-understood and agreed upon by the group. Then collectively and individually the team members can set out to address the causes, thereby reducing the number of resulting nonconformities. The team structure and composition is critical to its success. At a minimum, the team should include the operators, the adjudicators or inspectors, the supervisor of the operation, and a quality advisor. The quality advisor's role is to keep the team on track, advise on survey methodology as well as CQI, assist in the preparation of summary reports and data analyses, and act as a liaison between the team and higher management, if necessary.

The so-called CQI team for the operation may meet frequently (weekly or biweekly) when the operation is active to review the results from the time period since the last meeting. The primary objective of the meetings is to consider the most prevalent types of nonconformities as identified by the Pareto analysis and, using whatever data is available, discuss the possible causes and remedies. These discussions may lead to changes in the procedures, feedback to operations upstream regarding the quality of their outputs, retraining of the operators, changes in the work environment, and so on. In some cases, it may be determined that the process by which the preferred performance of the operation is determined is faulty. For example, there may be misconceptions among the adjudicators which lead to inaccuracies in the results of the inspections and false reports of nonconformities. These problems can be discovered in the CQI meetings and, in this way, both the original operation and the adjudication or inspection tasks can be improved.

Another objective of the CQI team is to assess the success of corrective measures which the team has implemented for the operation. Since the goal of CQI is a steady, continuous reduction in the overall nonconformity rate, the number of nonconformities in the operation should be closely monitored. Over time, there may be considerable variation in the types of nonconformities which are identified as most problematic by the Pareto analyses. Ideally, as the group focuses on and emphasizes improvement for a particular type of nonconformity, the frequency of that nonconformity should be reduced and some other type of nonconformity will rise to the fore. As these nonconformities are reduced, new classes will take their place, and so on. Over time, each type of nonconformity may take its turn in the top position while the overall nonconformity rate is ever decreasing.

Finally, the topics of the CQI meetings need not be limited to a discussion of the causes of the nonconformities or the group's progress toward reducing them. There may be other issues related to the work environment, shift structure, operator's manual, management practices, and so on that the team may discuss. The essential element in the meetings is open, uninhibited communication without fear of retribution. Creating this atmosphere is essential in order to fully understand the root causes of the nonconformities. It is a good practice to document the decisions of the group and distribute these to the group and possibly beyond.

Step 4. Implement the corrective measures and return to Step 1.

The measures to be taken to correct problems which give rise to the nonconformities may take a number of forms. For example, the individual operators may need to adhere more closely to procedure, now that these procedures have been clarified. The corrective measures, such as a change in procedure or the work environment, may be the responsibility of the facility manager or operation supervisor. Any changes to procedures, training, etc. should be well-documented.

In the next section, the results of an application of our approach to CQI to industry and occupation coding will be described. This application will illustrate in some detail how the CQI process can be put into practice and the potential benefits that can be derived from this strategy.

3. AN APPLICATION TO INDUSTRY AND OCCUPATION CODING

In this section we describe an experiment we conducted using CQI in our industry and occupation (I&O) coding operation. The experiment took place during the twelve months of 1992. However, due to its success, what began as an experiment has since become part of the standard operating procedures in RTI's I&O coding division. A general introduction to I&O coding at RTI, the procedures for initiating the CQI process, and the results of the year-long experiment are presented below.

3.1 The I&O Coding Quality Control Process

Questions used to obtain detailed information about industry and occupation are included in many surveys. For the most part the questions are open-ended; requiring the interviewer to record a verbatim response and to probe effectively until a complete answer has been obtained. For the information collected to be useful in statistical analyses, however, the verbatim responses must be coded using a standardized system of industry and occupation codes. In the United States these codes are developed by the Census Bureau and are updated after each Decennial Census to reflect new industrial and occupational areas. Currently, the system includes more than 30,000 occupation titles classified into approximately 500 occupations, and 20,000 industry titles classified into 230 industries.

At RTI, I&O coding is a manual operation. Specially trained coders are responsible for matching the open-ended survey responses to one three-digit industry code and one three-digit occupation code. The coding process may be described as follows: Upon receipt from the field, the questionnaire responses are keyed. A total of 99 characters are allowed for the industry response and an additional 99 characters for the occupation response. I&O coders work at terminals, accessing one data record at a time. Each record is coded independently by two coders. If the two coders are in agreement for both the occupation and the industry codes, the record is finalized with those two codes assigned. However, if one or both of the codes disagree between the two coders, the case is flagged for adjudication. I&O adjudication is handled by more experienced coding personnel. Each record sent to adjudication is reviewed. The codes assigned by each coder are displayed on the adjudicator's screen (though any information indicating which coder assigned the codes is not), and the adjudicator may assign one of those codes or a different code entirely. Regardless of whether the industry or occupation code is in disagreement, both codes will be reviewed during adjudication, as a change in one may result in a necessary change for the other. The codes assigned during adjudication are considered final codes and are written to the permanent data record.

Note that the RTI I&O Quality Control system, is a 100% (no sampling) inspection system. Every record in disagreement is referred to adjudication and every code is finalized only when two coders agree on the codes to assign or the adjudicator assigns the code. However, one should not conclude from this discussion that no error exists in these final codes. Two sources of error may still be present: erroneous agreements between coders, and errors made by the adjudicators. Both of these sources of error can be minimized by reducing the overall error in the system. To the extent that all coders are equally well-trained and capable of coding in accordance with the general rules, erroneous agreement between coders can be reduced (and theoretically eliminated entirely). Likewise, the fewer cases sent to adjudication, the fewer chances there are for adjudicator error.

The I&O coding procedure described above has been used at RTI for a number of years. The system has allowed us to fulfill our clients' expectations of obtaining high quality data. Yet, in 1991 we discovered that our quality control system was resulting in especially high costs for the I&O coding operation. These high costs were due to the fact that close to half (46%) of all cases were being sent to adjudication for final code assignment. The added cost was the result of the additional time billed by the adjudicators. With this level of disagreement between coders, it seemed clear that the two sources of error capable of infiltrating our system (erroneous agreement and adjudicator error) were likely to be a nontrivial source of error in the final codes.

In order to document the problem more fully, we developed a new measure of coding accuracy designated as the "coder error rate" (CER). The CER is calculated for each coder individually and is defined simply as the number of disagreements with the <u>final</u> code divided by the total number of codes assigned. We prefer to use the CER as our measure of accuracy rather than the disagreement rate because it allows us to classify the coding errors according to the adjudicator's code which we consider to be the most accurate code. In this way, we can identify industries and occupations which are particularly difficult to code. The CER will usually be significantly less than the between-coder disagreement rate -- almost half in many cases. Of course, lowering the CER will also result in a decrease in the between-coder disagreement rate.

Using the same 1991 data, we calculated the CER for every coder - for both industry and occupation. Overall, the error rate was 17.4 percent for industry and 21.1 percent for occupation.[3] However, there was a wide fluctuation of error rates among the coders. Industry error rates ranged from 13.0 percent to 28.1 percent. Occupation error rates ranged from 15.8 percent to 33.2 percent. While it is difficult to know what an acceptable error rate "should" be, discussions with a member of the Quality Assurance Staff at the Census Bureau indicated the Bureau averages error rates of 13.0 percent for industry and 18.9 percent for occupation.[4] Based on this information we felt certain that our error rates could be decreased and costs reduced. Using the 1991 error rates as a starting point, in January 1992 we began to implement changes in our I&O coding operation.

3.2 The Effects of System Changes on the Error Rate: Quarters 1 and 2

During the first quarter of 1992, a few significant changes were made to the I&O coding system. First, I&O coding was restricted to the day shift. Due to insufficient supervision and less motivated personnel, the error rates for night shift operators were consistently higher than those of the day shift. Schedules of work were reassigned so that all specialized coding work could be completed by day shift operators. Also, based on comments received form the coders, an enhancement was made to the on-line coding system. Rather than simply accepting a three-digit code keyed by the operator, the computerized system was reprogrammed to first display a written description of the code keyed by the operator. After reviewing the description, the operator could either enter the code or correct it by inserting a different code. All operators were in agreement that this change would improve their coding accuracy particularly in the area of catching simple "typos" before they were entered into the system. Finally, all operators were encouraged to work to improve the quality of their work. At the start of the second quarter of 1992, we reviewed the results from Quarter 1 with all coding staff, and continued to stress the importance of quality in the operation.

The results from Quarters 1, 2, 3, and 4 are presented in Figure 3.1. The improvement from 1991 to the first quarter of 1992 are clear. The CER for industry fell from 17 percent during 1991 to 8 percent by the end of the first quarter. Likewise, the error rate for occupations decreased from 21 to 12 percent. While these are dramatic improvements, our success was tempered by the fact that after an initial drop in both error rates, there appeared to be no additional improvement throughout the first quarter period. Statistical analyses show that the slopes of the lines displayed in Figure 3.1 do not differ significantly from zero. Thus, while an improvement was achieved, it was clearly not the continuous quality improvement described in Section 2 of this paper.

The error rate for industry decreased to approximately 7 percent by the end of Quarter 1. During the same period, the occupation error rate fell to just over 10 percent. From the graphics in Figure 3.1, it appears that some improvement may have occurred during this quarter. However, due to the fluctuation in the error rates throughout the quarter, our analyses once again showed no statistically significant change occurring during the quarter. Thus by the end of Quarter 2, while we had dramatically improved our overall error rates compared to 1991 data, we still had not been able to create an environment of continuous quality improvement.

3.3 The Effect of CQI on the Error Rates: Quarters 3 and 4

Prior to the start of Quarter 3, we implemented the full four-step CQI process outlined in Section 2 of this paper. Our first step was to make one simple modification to the RTI I&O coding Quality control System. The addition of the feedback loop allowed coders to receive information about cases they had coded incorrectly and to use this information to improve their future performance. More detailed information regarding the type of feedback given to coders and the way in which this took place is provided below.

To begin, weekly quality circle meetings were organized. All coders and adjudicators, as well as supervisory staff and a quality advisor took part in these meetings. During these meetings the coding staff was encouraged to share any problems they were encountering as they completed their work and to discuss possible solutions. Pareto charts were provided to the coders the day before the meeting. These charts documented the most-often misassigned codes for the group as a whole, based on all cases coded during the past seven days. This allowed the coders to see exactly which codes were causing difficulty for the group. Five industry codes and five occupation codes were documented in this way. In addition to the overall charts, each coder also received an individual listing which showed the ranking of these same problematic codes for his/her work alone.

Further information about these problematic codes was provided to the coders in the form of a Personal Errors Listing Sheet. For each of the codes identified in the Pareto analyses, coders received a listing of up to five cases which they had not coded correctly and had thus been sent to adjudication. This listing displays the entire text of the response as the coder originally viewed it. The listing also shows the incorrect code assigned by the coder, the code assigned by the other coder, and any comments made by the adjudicator. During the meetings coders were able to look at these examples and discuss how they had arrived at the incorrect code. Supervisors could then provide explanations and retraining to reduce misunderstandings about the codes and to increase the likelihood that these codes would be used correctly in the future. The adjudicators could also provide their rationale for assigning a particular code to a case.

The critical component of these weekly meetings was the use of a team approach. The Pareto analyses identified the most problematic codes for the coders as a group. The goal was not to identify poor coders and replace them, but to have all the coders focus on the most important issues for the group. By having each member of the coding staff strive to improve his/her work on the problematic codes for the week, the overall error rates decreased and individual performance improved as well.

The weekly meetings were not restricted solely to discussion of the Pareto analyses and personal listings, however. Coders and adjudicators were encouraged to bring up problematic issues related to their work environment, the quality of the data they worked with, and other demands on their time which impinged on their ability to work efficiently. The quality advisor who attended each meeting would, when necessary, act as a liaison to upper management and staff in other divisions of RTI whose decisions impacted on the coding operation. In this way, the coding operation could be improved both by increasing the skill-level of the coders and by improving the external environment in which the coding operation occurs.

By the end of Quarter 4 the error rate for industry had fallen to four percent and the rate for occupation had decreased to just under five percent (See Figure 3.1). Not only is the overall decrease from the end of Quarter 2 to the end of Quarter 4 dramatic, but the slope of the lines for both industry and occupation during Quarters 3 and 4 show a statistically significant downward trend. Our goal of continuous quality improvement has clearly been realized.

An important question to raise at this point is, how much did the reduction in error rates cost to the coding operation. To address this issue, we first looked at the effect of the reduced error rate on coder production rates. During Quarter 1, production rates increased markedly from approximately 40 codes assigned per hour to just over 80 codes per hour. In Quarter 2 there is virtually no additional improvement, though the rate of 80 codes per hour is clearly sustained. Beginning with Quarter 3, when our efforts at adapting the full CQI model began, production rates dropped to about 65 codes per hour and remained fairly stable at this rate throughout the quarter. In Quarter 4, however, the production rate began to increase such that by the end of 1992 our coders were assigning nearly 80 codes per hour. At this rate, our coding operation has rebounded to nearly the same rate that was achieved prior to implementing CQI but with half the level of error present in the operation. We believe that the initial decline in production during Quarter 3 is entirely attributable to our CQI efforts. Coders began to concentrate more fully on the task and to take more time in looking up and assigning codes. But, as they received feedback during the weekly meetings and began to understand where and why errors were occurring they were able to combine speed with accuracy more effectively.

It is clear from these results that production rates returned to the levels attained prior to implementing CQI. However, this is not to say the costs to the operation were exactly the same as those incurred prior to adopting CQI. The lower error rates mean that fewer cases are in disagreement between the two coders. Thus, fewer cases must be sent to adjudication which reduces the cost of coding a case. However, the CQI methodology is not without added costs. There is the cost of the weekly meetings which involve the entire coding staff as well as the quality advisor. There are also costs associated with producing the Pareto analyses and costs for the quality advisor to handle issues that involve staff outside the coding unit.

In order to compare the costs associated with the traditional and CQI approaches we developed a simple cost model which take into account two types of variable costs: the cost of the coding and the cost of adjudication. The model for the average cost per code (CPC) is: $CPC = \frac{CHC}{CPH} + p \cdot \frac{AHC}{APH}$ where CHC is the average "coding hourly cost", CPH is the average "number of codes assigned per hour", AHC is the average "adjudicating hourly cost," APH is the average "number of adjudications completed per hour," and p is the estimated probability that a code is sent to adjudication. Thus, CPC decreases as coder productivity (CPH) and adjudicator productivity (APH) increases and as the proportion of cases being sent to adjudication, p, decreases. Figure 3.2 is a graph of CPC for the four quarters of the study. Despite the fact that the average coder production rate under CQI in Quarters 3 and 4 was lower, the average cost per code for Quarters 3 and 4 is almost identical to the average for Quarters 1 and 2 under the traditional inspection approach. Further, a clear downward trend is exhibited in Quarter 4 which inspires hope that even greater cost efficiency may be realized under CQI. This dramatic reduction in cost is somewhat unexpected since virtually no emphasis was placed on coder productivity in this experiment. With appropriate control and feedback of coder productivity in future implementations of the CQI coding operation, we expect the downward trend in CPC to continue with no increase in coder error rates. These costs savings are substantial considering the thousands of cases which are coded in the operation each quarter and more than offset the additional costs of implementing CQI that were noted above.

Finally, at the end of Quarter 4 we asked the coding staff to critique the CQI process. Opinions were unanimously positive. The staff felt the weekly meetings provided a nonthreatening environment in which they could raise questions and concerns. The meetings also allowed the coders to receive additional training in how to assign some of the more problematic codes. The meetings were viewed as an excellent forum for this type of retraining since all coders and adjudicators were present and thus there was no danger of some staff members failing to obtain the information. The Pareto charts and error listings were viewed as an important component of the weekly meetings. Coders felt the retraining was more practical because there were "true life" examples from which to work. All staff members felt the focus on group improvement was especially useful. The coders reported that they felt increasingly comfortable reporting problems to their supervisors or to other coders because the goals of CQI were based on the group working together rather than each coder working individually. The benefit of having a quality advisor involved in the CQI process were also noted. The quality advisor was able to impact change in other areas of the survey process which were beyond the "jurisdiction" of the coding staff.[5] This is especially important as the coding operation will clearly be affected by decisions made during an earlier stage of the survey process.

4. OTHER AREAS OF APPLICATION AND CONCLUSIONS

The coding example provides a useful model for implementing CQI in other survey operations. Applications of the four step approach to data entry, editing, and other coding operations are readily apparent. One factor that these operations have in common which we feel is key to the success of CQI is the ability of the operators to perform a step-by-step review of their actions in creating the output of the operation and to discuss, in a group setting, how their actions deviated from the preferred action. As mentioned in Section 3, the personal error listings that were provided to the coders prior to each CQI meeting were critical to the success observed for that operation.

This same approach may be used for editing. In editing, the operators (or editors) review the paper questionnaire for nonconformities in the output of the data collection operation that could pose problems for data entry or other subsequent operations. As an example, editors may fail to assign a code for a "refused" or "don't know" or may use an incorrect code. When this occurs the mistake may not be detected until the document is being keyed, and possibly not until the data is analyzed. Such nonconformities in the editing operation can be reviewed by the editors using an approach similar to the personal error listing for coders. That is, the editors can observe the input they originally received and can review the steps they performed in obtaining the output. In this way, CQI teams can constructively discuss the root causes for the editing nonconformities. Likewise, for data entry the input received by the operators can be reviewed, as can the actions taken to enter the data, and the resulting output. Thus, the personal error listings approach is possible since the actions of the operators can be unambiguously reconstructed and examined for possible root causes.

One set of operations that does not neatly conform to the I&O coding model is interviewing. Consider centralized telephone interviewing as an example. Centralized telephone interviewing consists of many interactions and interchanges between the interviewer and the respondent. To obtain a response (output) for a single questionnaire item, the interviewer may deliver the question, clarify the question for the respondent, probe to obtain an acceptable response, enter the response into the data collection system, and provide feedback to the respondent. Thus, to obtain a single output may involve a series of inputs and interviewer actions. Unfortunately, none of these inputs and actions is preserved in a form that would allow subsequent objective evaluation in a team setting as described for I&O coding. Without the equivalent of a "personal error listing" for interviewing, examining the root causes of nonconformities in the interviewing process, particularly with regard to interviewer performance, is indeed challenging. Tape recording (both video and audio) of live interviews would provide the needed data; however, tape recording is illegal in the U.S. unless the respondent is fully aware that it is being done. Therefore, the routine tape recording of interviews that would be needed for CQI is seldom done for fear of its damaging effects on respondent cooperation rates.

In the absence of the ability to examine, post-hoc, the inputs, actions, and outputs associated with interviewing, the advantages of recreating for the operators the process which lead to a particular nonconformity are lost. Rather the interviewers must rely on a recounting by a trained monitor (inspector) of the series of inputs and actions which produced one or more nonconformities. We find this to be a much less effective device for CQI for several reasons. The process of monitoring, especially using the system of Couper, et al. (1991) which a number of survey organizations have adopted, is itself quite prone to error. The procedure requires the monitor to assess each input and interviewer action associated with a particular response and to record their evaluations of these relative to preferred inputs and actions during a live interview. These assessments, which are made at the individual question level, include determining whether deviations from the written question changed the meaning of the question; whether a probe was used appropriately, completely and nondirectively; whether feedback to the respondent was appropriate and neutral; and so on. Our experience with this system indicates that monitors may be quite inconsistent in making these assessments in live interviewing situations. (However, monitoring consistency improves somewhat for tape interviews for which monitors are allowed to replay the respondent-interviewer interchanges.) As a result, these data have limited utility for CQI.

Further, unlike the case of I&O coding, in the interviewing operation, the operators are unable to conduct post-hoc evaluations of their own work. Rather, they are presented the assessments of the monitor, usually in an aggregated form, without any means of accurately tying individual nonconformities to specific events. Thus, the search for the root causes of the nonconformities is substantially impeded. We have experimented with providing feedback to an interviewer regarding the nonconformities observed in a particular monitoring session immediately following the monitoring session while the events are still recallable. However, this approach also is problematic. Part of the problem is the presence of errors in the monitor's assessments of the interview which was described earlier. Further, this form of feedback tends to focus the full responsibility for the nonconformities on the interviewer rather than the true root causes (which could include poor questionnaire construction, faulty computer hardware, or inadequate training).

For field interviewing, the problem of collecting data on the operation for CQI purposes is even more difficult. Performance measures such as item nonresponse, edit failures, and the results of reinterviews and interview verifications have been used to monitor field interviewer performance. However, these measures suffer from all of the same problems that were described for the use of monitor assessments for telephone interviewing. There is some evidence that the tape recording of field interviews may be quite feasible without any adverse affects on respondent cooperation rates (see Moore, et al., 1992). This would offer the possibility of post-hoc group and individual assessments of root causes.

Our future research efforts will be directed, in part, at implementing CQI in other survey operations, particularly for centralized telephone interviewing. Additionally, we are exploring extensions of the CQI approach by incorporating the ideas of statistical process control. In particular, we are currently investigating the use of process control charts (see, for example Wadsworth, et al., 1986) for identifying special causes which result in abnormally high numbers of nonconformities in an operation, relative to historical data. As an example, operators who have nonconformities in their work assignments which tend to be much larger than the group mean may be identified using control charts. Then, corrective measures may be directed toward the root causes which are specific to the operators. However, there are real risks in this approach. In this paper, we have emphasized a team approach to reducing the number of nonconformities. In operations where the team approach has been implemented, the operators have commented they enjoy the group approach and do not feel threatened or unfairly judged by it. To now focus on the individual operator as an assignable cause could be received quite negatively by the operators and thus adversely affect the morale of the group. Further, it remains to be seen how much additional improvement is realized by adding individual-targeted corrective measures to the team approach relative to the team-only approach. Other uses of control charts in survey operations will also be investigated.

Endnotes

[1]This is the simplest form of acceptance sampling. More complex sampling schemes are often encountered in practice. For a description of alternative acceptance sampling schemes, see, for example, Wadsworth, et al. (1986).

[2]For more information on Pareto analysis, see, for example, Wadsworth, et al., 1986.

[3]It is typical for occupation error rates to be higher than those for industry as the responses are usually more difficult to code and also because the codes are driven from the industry code which is assigned first. Thus, if the industry code is assigned incorrectly, it is more likely that the occupation codes will also be in error.

[4]Personal communication with Phil Gbur, Quality Assurance Staff, U.S. Bureau of the Census.

[5]Examples of tasks undertaken by the quality advisor include: 1) development of an improved interviewer training module on how to collect sufficiently detailed I&O data, 2) enhancements to the on-line coding software, and 3) development of a mechanism to notify interviewers who are not collecting sufficiently detailed I&O data.

REFERENCES

Couper, M.P., Hollard, L., and Groves, R.M. (1991). "Developing Systematic Procedures for Monitoring in a Centralized Telephone Facility," *Journal of Official Statistics*, Vol 8, No. 1, pp. 63-76.

Deming, W.E. (1986). *Out of the Crisis*, MIT Press, Cambridge, Mass.

Ishikawa, K. (1990). *Introduction to Quality Control*, 3rd Edition, JOSE Press LTD, Tokyo, Japan (reprinted in English by 3A Corporation)

Moore, J.C., Bogen, K.E., and Marquis, K.H. (1992). "A Cognitive Redesign of the Survey of Income and Program Participation: Development of Procedures and Initial Test Results," Paper presented to the Census Advisory Committee of the American Statistical Association, April, 9-10, Suitland, MD.

Wadsworth, H.M. Stephens, K.S., and Godfrey, A.B. (1986). *Modern Methods for Quality Control and Improvement*, John Wiley & Sons, NY.