QUALITY IMPROVEMENT IN SURVEYS - A PROCESS PERSPECTIVE

Lars Lyberg, Lilli Japec, Statistics Sweden; Paul Biemer, RTI
Lars Lyberg, Statistics Sweden, R&D, P.O. Box 24300, S-104 51 Stockholm, Sweden

Key words: Quality, Survey processes, TQM

1. What is quality?

Quality has become one of the current buzzwords in society. The concept is vague and has different meanings in different contexts. In textbooks on management issues, as well as in advertisements, commercials and speeches one might see or hear slogans like "quality is here to stay," "quality for profit," "quality- a management commitment," and "when quality leads, productivity follows." These slogans are not definitions of quality but they reveal that quality not always can be taken for granted and that it is sometimes considered expensive and not worthwhile.

During the last 20 years we have witnessed a quality "revolution." Gurus like Deming, Juran, Taguchi, and others have been on the forefront advocating the need to better understand quality and how to improve it. Quality improvement implies change and there is a process for change, just as there are processes for car manufacturing and statistics production. Successful organizations know that continuous improvement is necessary to stay in business and they have developed measures that help them change. Typically such measures include a style of management that emphasizes quality, customer orientation, empowerment, teamwork, a scientific approach, understanding variation, decisions based on data rather than gut feeling, the use of specific management and statistical tools and evaluation of approaches according to, for instance, Deming's Plan-Do-Check- Act procedure.

How should we then define quality in general? There are easy answers like "fitness for use" (Juran and Gryna 1980) and "the totality of features and characteristics of a product or service that bear on its ability to satisfy a given need" (Swedish Standardization Committee). These easy answers quickly become complex when we realize that whenever there is a variety of uses (as is the case of statistical products) fitness for use must have multiple quality characteristics, where characteristic importance varies among users.

It is important to distinguish between two major quality characteristics, quality of design and quality of conformance. A booklet with colored diagrams of statistical data looks nicer than an excerpt from a database and is an example of design quality. Generally higher design quality costs more. Quality conformance is the degree to which the product conforms to the intended use. Typically conformance is a predetermined margin of the error of an estimate of a population parameter. Sometimes it is possible to increase this type of quality characteristic and reduce costs at the same time.

2. Quality of surveys

The quality of surveys is a mixture of design quality and conformance quality. Survey quality can be seen as a vector where components could include: error, level of confidentiality assured, wealth of detail, user access to data, and so on. Traditionally, though, survey quality is seen mainly as a function of the survey error or sometimes total survey error, i.e., quality is conformance quality. Various design quality characteristics have not always been part of an allocation process aiming for good total quality.

The total survey error can be modeled in various ways. One well-known approach is the US Bureau of the Census survey model (Hansen et al 1964) where the mean squared error (MSE), $E(y-X)^2$ or $E(y-Z)^2$, is decomposed into sampling variance + response variance + covariance + squared bias (relative to X or Z) and where y is the estimate of X, a parameter of a preferred process that is defined as the best we can do with unlimited resources or of Z, the ideal or true parameter, a value that is unknown and may be unmeasurable but is nonetheless the goal. The MSE is a very useful tool for determining how to allocate resources in a survey so that they will do most good. For example should we put more money in nonresponse follow-up or interviewer training. Using the MSE as a guide, we can begin to address the cost-error tradeoffs. There are a number of problems with associating survey quality with the MSE only:

The mean squared error, i.e., $E(y-Z)^2$ is one measure of survey quality. Unfortunately, in most surveys some of the MSE components cannot be estimated, not even with an abundance of resources available, and Z, will differ from the expected value of our actual estimate y.

If ambitions are lowered and Z is replaced by X, we are still in a difficult position. $E(y-X)^2$ is now a measure of the quality. Most survey organizations can only provide underestimates of this quantity, due to lack of financial and methodological resources. A good estimation would require a set of preferred and more expensive methods and we would still be left with uncertainty not accounted for.

Even if we should have access to estimates of the total error or a similar quantity its value is relatively limited. Such attempts lead to estimates of post-survey quality measures or quality indicators. The estimates are important for deciding data accuracy, but, except for repeated surveys, may be of limited value for improving the survey data.

Before any meaningful release of measures of survey errors can be made one has to administer some kind of quality control procedure for the most important error sources, for instance for the listing procedure, the data collection, the coding and the data entry. This usually calls for verification procedures, sometimes on a mass-inspection basis. The use of inspection may be criticized on a number of points (see Biemer and Caspar 1994) but basically it is costly and does not generate any continuous improvement.

As mentioned MSE or other conformance measures normally do not take design quality components into account.

Many survey organizations throughout the world are now working with the concepts of Total Quality Management (TQM) in the context of design, data collection and data processing. The basic idea is that interest must shift from mass inspection or verification and post-survey evaluation to controlling the survey processes, because product quality is achieved through process improvement. By concentrating on process improvement it is possible to address improvement issues regarding both design and conformance quality. Methods for process quality improvement such as use of TQM tools, team work, identifying customer needs, understanding and reacting properly to different kinds of variation and a management style advocating continuous improvement are now being successfully applied in these organizations.

Statistical organizations should be especially fit to work along these lines (Colledge and March 1993) since staff members should be used to properly collect and analyze data. Statistical organizations should know how to conduct studies and experiments and know the value of having access to good data that could benefit sound decision-making. Unfortunately this is not always the case, though.

3. Survey Processes and Quality

We have seen that there are problems with traditional measurement and control of survey quality as a function of errors. Measuring errors has its advantages when establishing error rates and error structures so that proper allocations of resources can be made for future error reduction efforts, for instance by ranking error problems according to some Pareto principle. Such evaluation studies are large-scale by nature and cannot be a regular part of every survey, since most customers would be unwilling to fund them. Evaluation studies are consequently relatively rare. Good examples are the studies conducted as part of the U.S. Censuses of Population and Housing (see for example U.S. Bureau of the Census 1972). Results from those studies have led to changes in subsequent censuses regarding choice of data collection mode, coding system and design of census forms.

In ongoing surveys and as an attempt to "control" the error various forms of verification and inspection are common. The purpose of these systems is that a prespecified outgoing average quality limit should be assured, which calls for recoding, reprinting, rekeying, etc. of material that is classified as being too erroneous. The theory for this approach was developed by Dodge and Romig (1944) and was originally used in the production of war material during World War II. The basic idea is that substandard work discovered in a sampling inspection process is scrapped and replaced by error-free items. Later this theory was adjusted and used in administrative applications like statistics production and in operations like coding and data entry (see Minton 1972 and Lyberg 1981).

As pointed out in Deming 1986 and Biemer and Caspar 1994, inspection is administered because the underlying process is error-prone. Thus an alternative to inspection is to try to make the underlying process virtually free from errors. If the process is free from errors so is the product. Inspection has many drawbacks besides being expensive. The inspection process generates errors of the first and second kind, which lead to shifts in the operating characteristic curve. Having special "control departments" taking care of the inspection makes operators take less responsibility for the quality of their work resulting in a lack of motivation to improve one's own work (since it is the responsibility of the control department). Feedback to operators based on rejected work lots sends the wrong message. There is a general rule of thumb that says that of all errors in systems those operating the systems are responsible for 15 % of the errors while the rest is caused by deficiencies in the systems themselves. To put the blame on operators for all errors is therefore truly demoralizing, especially if feedback attempts do not discuss potential root causes of the procedures.

Thus a natural route for survey quality to take is to move from inspection to a state with continuous process and product improvement where the ultimate goal is to achieve the smallest error rate possible.

4. The Process Perspective

The process perspective in surveys is discussed in, for instance, Morganstein and Hansen 1990 and Morganstein and Marker 1997. The more general process perspective is discussed in Scholtes 1994. As noted above, each error source can be seen as one or several processes. Let us assume that the survey questionnaire is one such error source. Typical errors resulting from questionnaires include response errors, interviewer errors, data entry errors and nonresponse. To measure and control these errors one might have to administer evaluation programs, reinterviews, measurements of the correlated response variance, editing, nonresponse follow-up, and imputation. Needless to say, these efforts are costly and time-consuming. When trying to follow Deming's and similar philosophies, we should try to build quality into the processes to prevent errors rather than identify errors once they have occurred and replace them with more accurate data. This building of quality is not done overnight. It takes some continuous effort.

As pointed out in for instance Batcher and Scheuren (1997) automation efforts like modern CASIC approaches make it easier to perform measurements of key process variables. With such approaches it is possible to tabulate interviewer results and perform statistical monitoring on a daily basis. Specific interviewer results regarding refusal rates, unusual responses, completion times, time for contact attempts, number of contact attempts can be quickly identified as can possible root causes to problems. Data on keying can reveal where in the questionnaire interviewers had problems, for instance where they were forced to turn back to previous questions.

In Morganstein and Marker 1997 the general elements of a plan for continuous improvement are outlined. *The first step* is to identify the critical product characteristics. These are mainly determined by customers' needs and expectations and are mixtures of design and conformance quality characteristics. In our questionnaire example there might be broad needs shared by many customers like specific combinations of survey topics and data collection modes or more specific customer needs regarding, for instance, a large number of questions, the inclusion of sensitive questions or that CATI should be used, no matter what. Other common and critical elements at this stage include time schedules and budgets, which have a bearing on the choice of mode and the number of questions possible and whether the customer wants to carry out some questionnaire work on his or her own. Some organizations carry out continuing customer satisfaction surveys to find out about general and specific needs. Such surveys tend to be rather general and it is hard to see how a customer satisfaction survey might be efficient for shedding some light on the questionnaire issue, as compared to discussions with the customers.

The second step is to understand the process. Here we can use one of the TQM-tools, the flow-chart. By means of the flow-chart all processes related to the development of the questionnaire are visualized. The process details are worked out by a team consisting of representatives of those actually working with the processes. In our case the members could be, say, interviewers, cognitive researchers, survey managers, the users and people working with data entry. Our experience is that the team effort is essential in this step, since often there are divergent opinions on process details among those involved.

The flow-chart should identify process details, flows and sequences, distinguish between activities and decision-points, identify process owners and responsibilities and when possible identify customers for the different steps. The flow-charting should result in a consensus and a common understanding of the processes. Once consensus has been obtained obvious problems can be identified and fixed. Such problems include both glaring problems and more subtle inconsistencies that are due to lack of documentation and inadequate knowledge about best practices available. Then it is possible to streamline the processes by eliminating errors and trimming fat.

The third step is to identify key process variables, i.e., factors that have a large effect on product characteristics. If a small item nonresponse rate is an important product characteristic in our example, then factors like question wording, order of response alternatives, placement of sensitive questions and questionnaire length might be vital. If the impact of factors can be described by frequency data then a Pareto diagram can help distinguish between "the vital few and the trivial many." When frequency data are not available then fishbone or cause-and-effect diagrams can be used. In such diagrams all possible factors affecting the desired outcome (a small item nonresponse rate) are listed and the improvement team chooses those it believes have the largest effect on the outcome and these are the factors to measure first.

Thus measurement is vital. Without measurements it is not possible to say whether a process is improved by a specific approach. At the very least, it is not possible to say to what extent the process has improved.

The fourth step is to evaluate measurement capability. We have just said that measurement is vital but equally important is that data collected for quality improvement are themselves of good quality. Statistical organizations should be especially well suited for this task, since it is their job to collect meaningful data. Nevertheless, there is evidence that also in statistical organizations process measurements can result in data that are not capable of supporting quality improvement work. One example of such data are those emanating from some customer satisfaction surveys, where, for example, effects of question scales and nonresponse can limit their use for improving quality.

The fifth step is the reduction of variability. The key process variables that have been identified should be tested for stability. The goal of this step is to make the processes predictable or stable. If one cannot measure relatively precisely one will be chasing ghosts in the improvement work. There are two types of variation in this context: special cause and common cause. A stable process is one where special causes are virtually eliminated.

The variation is visualized by means of control charts where measurements of key process variables are plotted. Associated with the control chart are an upper bound and a lower bound, which are usually set at +/- 3 standard errors from the overall process average. The bounds are established after a number of measurements have been made so that they are based on real data rather than specifications. The process should be monitored until there are no signs of special cause variation. Virtually all variation should be of the common type where all plotted measurements should fall inside the bounds of the control chart.

It is very common that managers react to common variation as if it were special, thereby the term "chasing ghosts." This is a costly and frustrating approach resulting from not using proper measurements and control charts.

The sixth step is to establish a system for continuous monitoring of processes. Suppose the process is stable after elimination of the special causes one by one. There is only common cause variation left and it is time to improve the system. Remember that we were looking for processes designed so that quality was built in. Most of the time improvement means narrowing the control limits if there is dissatisfaction with the kind of variation that the current stable system generates. It can also mean shifting the process average to a higher or lower level. For this to happen the system has to change.

So how do we create a process for questionnaire design where quality is built in rather than having a situation where errors have to be evaluated afterwards, or where errors and other problems have to be fixed during the data collection? In a continuing survey this could be done by having the team work on the Plan-Do-Check-Act cycle. Based on information available from flowcharts, fishbone diagrams, Pareto diagrams and other sources a number of ideas for improvement might come up. The most promising idea is planned as a change or a test (P). The change is conducted, preferably on a small scale (D). The results are checked. What did we learn? (C) Based on the outcome we decide whether the change of the process should be abandoned or adopted. A third alternative is to run it through the cycle again and this time under different experimental conditions (A). If, after adopting the change, there is need for further reductions of variation the other ideas standing in line can be put to a test.

Let us say that we have problems with item nonresponse in our continuing survey. Based on what we know we may try changes regarding question wording, placement of sensitive questions, scales, number of response alternatives, and/or changes regarding questionnaire layout to make use of modern navigation principles. The change(s) are tried according to the P-D-C-A cycle and are abandoned or adopted depending on its effects on item nonresponse rate and its consequences on other product characteristics like total cost and measurement error.

In a statistical organization the development of questionnaires is a recurring process. Some surveys are conducted just once and P-D-C-A would have no meaning unless there are time and resources for extensive pretesting. An alternative version of the P-D-C-A cycle could the be S-D-C-A where S stands for Standardize. This is appropriate when the goal is to improve methods used on a process. An obvious way to decrease variation in such recurring processes is to have all employees working on a process use the same methods and procedures. There are a number of such recurring processes in a statistical organization, for example questionnaire design, nonresponse rate reduction, nonresponse weighting, editing, imputation, parameter estimation, variance estimation, coding, data entry, analysis and so on and so forth. A reasonable strategy for a statistical organization would be to develop Current Best Methods (CBM) for its major recurring processes, to have them implemented and continuously updated as new knowledge is generated.

As pointed out in Schwarz (1997) the general process of questionnaire development could be improved considerably. There is a very extensive literature on pretesting, use of cognitive methods, results from experiments on question wording, question order, and questionnaire length, data collection mode and questionnaire design. Despite this abundance of knowledge there is extensive variation within agencies as to how this knowledge is used. Typically approaches range from fairly continuing improvement efforts to just patchwork or extremely infrequent testing of new ideas. The variation between agencies and countries is of course also extensive, which is starting to create problems in the increasingly popular international surveys on, for instance, literacy.

A system change in the area of questionnaire design would involve the development of a CBM that could be used throughout the organization. That is probably the most effective way to improve quality, to let everybody use a standardized process where the methods used on the process are the best we currently can think of.

In questionnaire design such a process would include activities like definition of research objectives, translation of research objectives into survey questions, writing questions, pretesting, question asking, question answering (understanding, recalling, formatting, editing and entering responses), layout and navigation, and data collection mode issues.

5. Current Best Methods

In Morganstein and Marker (1997) the role of CBMs in the improvement of survey quality is discussed in detail. They state that one of the most frequently identified sources of variation is the difference in performance or even approach among people assigned to do the same task. Above we have provided examples of processes that are so common and vital that a standardization of them would be central to improved quality.

At Statistics Sweden two CBMs have been developed, one on nonresponse rate reduction (Japec et al 1997) and one on editing (Andersson et al 1997). Here we intend to describe the development of the first of these. Several others are underway.

To keep nonresponse rates in surveys conducted by Statistics Sweden on decent levels is the responsibility of individual survey managers. During recent decades nonresponse rates have increased in many surveys or been stable on high levels. A few years ago top management decided that efforts should be made to improve the situation. It was obvious that levels of ambition and knowledge of good practices varied across surveys resulting in less efficient processes for nonresponse rate reduction. It was decided that a CBM should be developed with the intent of reducing variation in performance and consequently a reduction in final nonresponse rates. Morganstein and Marker (1997) provide a flowchart for the development of CBMs. The first box in that chart is the identification of vital processes. We interpreted that as collecting data on practices used in the surveys.

For the last 13 years Statistics Sweden has collected and plotted nonresponse rates for a number of its surveys. The document where these rates are published is called the Nonresponse Barometer. The number of surveys included in the Barometer has gradually increased over the years and it now comprises about 50. The Barometer does not, however, contain much information on methodology used for reducing these rates. Therefore it was decided that a survey should be conducted to shed some light on how the vital processes, like the contents of advance letters, data collection strategies, follow-up, questionnaire design, interviewer training, respondent burden, and the use of incentives were dealt with. Initially it was assumed that these descriptions could lead to some kind of Pareto analysis where some crucial process steps were identified and emphasized in the forthcoming CBM. It turned out that not a single one of these process steps were handled satisfactorily by a majority of the surveys in the study.

Given all the work that has been laid down on nonresponse rate reduction over the years the survey showed that surprisingly little of what is known about methodology had been picked up by the survey managers. Furthermore there was a general lack of data on nonresponse and nonrespondents that could guide survey managers in their improvement work. Very few had put in great efforts in developing advance letters. Only minor revisions had been made during the period 1990-95 and only one study on the effects of advance letter contents had been conducted during the period. The advance letters sent out to respondents of mandatory surveys were often impolite and demanding. The mandatory feature is sometimes used as a threat. The layout of the letters was often gray and dull and the language dry and bureaucratic. Some contained grammatical errors. Needless to say some letters looked nice and were quite informal.

Many of the data collection processes seemed to be old, and when in need for improvement, had been changed without any P-D-C-A approach. The time for sending out reminders and the number of reminders had, for a vast majority of surveys, not been based on data on response in flows. We found that for similar surveys using the same main data collection mode the timetables for reminders and the number of reminders varied a lot. It was evident that some data collection processes were unnecessarily dragged out with periods of very low activity. Some surveys had established nonresponse goals. One example of such a goal is that the nonresponse rate should not exceed 20%. Some of these goals had been decided by the customer but in other cases it was unclear how they had been decided. Some of the business surveys used selective nonresponse follow-up but no survey of individuals and households did that, thereby basically leaving it to the interviewer to choose which respondents to follow up on.

As for questionnaire design, a vast majority of the surveys had not used Statistics Sweden's cognitive laboratory for testing their questionnaires. As a matter of fact only one of the questionnaires sent to establishments had been tested by the laboratory. Furthermore there seemed to be a variation in opinion among survey managers whether certain questions are sensitive or not. No special methods for dealing with sensitive questions were used in any survey.

Also views on what should be considered a heavy respondent burden vary between survey managers. Some of the surveys had an extremely heavy respondent burden by any standard. Measures had been taken in some surveys to deal with this problem.

Incentives were used in some surveys but it did not seem as if survey managers were aware of recent research on how to administer incentives.

The interviewers made their own decisions when it came to allocating contact attempts over time. Naturally most contact attempts took place during hours that were convenient to the interviewers. There is a clear tendency that the number of contact attempts decreases over the week with virtually no attempts being made during the weekend. This is in contrast with data showing that people are more available for interviews during evenings and weekends. Strategies for tracking respondents and converting refusals were usually worked out by the Interviewing Unit and survey managers had little knowledge about these processes and how interviewers were trained in them.

Armed with these rather discouraging results we started planning the CBM. It was quite clear that the text could not be a cookbook containing specific recipes. Surveys are different and it is not possible to do things exactly the same way in different surveys. Rather the CBM should provide a TQM framework for improvement work by emphasizing the use of known dependable methods and a general encouragement to define key process variables and collect process data.

The CBM was developed by a team consisting of three people from the R&D Unit, two statisticians from each of two subject matter departments and one behavioral scientist from the Interview Unit. Just by coincidence five of the team members were trained TQM project facilitators (one is usually sufficient). To begin with the team was much larger but that structure was soon abandoned due to communication and responsibility problems. A suitable team size is 6 to 8 people. The team started out conducting the above-mentioned survey and the associated analysis of the problem picture. Then the structure of the book was decided and the general contents agreed upon. A lot of time was devoted to searching the literature for known dependable methods and to do benchmarking at other agencies. Chapter texts were drafted and reviewed by a group comprising about fifteen people from various parts of the organization. Top management took a deep interest in the development and as a result a very large group was involved in various stages of the work, thereby making implementation easier.

The resulting CBM is a book consisting of four parts. The first part deals with basic notions like definitions and calculations of nonresponse rates, reasons and categories of nonresponse and recent theories of survey participation (Groves and Couper 1998 and Dillman 1978). The second part concentrates on what is called the main processes, those that are present in virtually all surveys. Processes dealt with include questionnaire design, advance letter design, follow-up procedures, privacy and confidentiality assurance, data collection and how to combine various measures to achieve decent response rates. The third part concentrates on processes that are not always part of the survey, like dealing with sensitive questions, respondent burden, interviewer issues, using and administering incentives, using proxy respondents and panel surveys. The last part provides a framework for identifying and measuring key process variables so that each survey manager can lead his or her own improvement work. Examples of such key process variables are nonresponse rate by sample breakdowns, nonresponse rate by collection mode, tracing source hit rate, average number of contact attempts, distribution of contact attempts over time, inflow by reminder waves, refusal conversion rate, cost for collecting data on the last 10% of the respondents and item nonresponse rate per variable.

The book has been widely disseminated and used in local TQM-projects. Much implementation work is still needed, though, since its use is far from uniformly spread across the agency.

6. Some Experiences from Research Triangle Institute (RTI)

In Biemer and Caspar (1994) continuous quality improvement of survey operations is discussed as an alternative to inspection methodologies. The ultimate goal of continuous quality improvement is to change the actual performance of an operation or a process so that the number of nonconformities (the difference between actual and preferred performance) in the operation is reduced over time to zero. RTI's approach is an adaptation of the P-D-C-A cycle and could be described in the following way:

Step 1. Perform the operation and observe the nonconformities.

By applying preferred procedures on a sampling basis it is possible to observe the nonconformities. Nonconformities in interview work can be established by a supervisor or by using a system for reinterviewing. In coding the same can be accomplished by an expert coder or by using a system for independent verification with adjudication of differences.

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

The, say, five most frequent nonconformities are dealt with in descending order. There might be a problem establishing type classes for different nonconformities. For interviewing there are systems of monitoring based on behavioral codes (for instance Couper, Holland, and Groves 1991) that can be used. For other operations some development of a minor classification system might be called for.

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

It is the job of a team consisting of people who work on the actual operations or processes to investigate the nonconformities until their root causes are understood and consensus has been reached. The team work proceeds according to principles outlined earlier in this paper.

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

As in the P-D-C-A cycle changes that the team agree upon are implemented and results of the process are closely monitored.

In an application to industry and occupation coding a number of system changes were made based on the team's analysis and data collection on nonconformities. Coding was restricted to the day shift and the number of coders was reduced from 15 to 5. These changes were made because of data on costs and error rates per shift. Another change concerned the on-line coding system. As soon as a coder had keyed his or her three-digit code a written description of the code was displayed on the screen. That gave the coder an opportunity to review the description and a chance to change the code assigned. This change was especially efficient for discovering typographical errors. Other changes implemented to generate continuous quality improvement included the addition of a feedback loop that allowed coders to receive information about cases that they had coded incorrectly so that future performance could be improved. Weekly quality circle meetings were held where all the coding staff participated. Pareto charts were provided to the coders showing the most erroneous codes for the group. Each coder also received individual listings of problematic codes as well as adjudication results on an individual basis. During the meetings coders could discuss the examples in the listings and how they arrived at the wrong codes. Supervisors could explain and retrain if necessary.

This approach went on for about a year and the results were dramatic. Error rates for industry fell from 17% to 4% and for occupation from 21% to 5%. Also costs went down from about $0.75 per code to about $0.42 per code despite the fact that no emphasis was placed on coder productivity in this application.

7. Cost issues

In the beginning of this paper we mentioned that sometimes it is possible to increase quality conformance and reduce costs at the same time. This is exactly what happened in the RTI example discussed above. Cases like these should be ideal ones for promoting ideas on continuous improvement, but resource issues seem to be a roadblock in many organizations. Typical concerns among managers are associated with prevention costs. Planning for quality on an organizational or product level, developing CBMs, training of staff, improvement project work, and preparation of quality reports are examples of such costs. These costs may seem high at the outset and could become a barrier to improving quality. It is therefore important to also look at all the failure and appraisal costs. Such cost figures are seldom secured in statistical organizations. Examples of costs are inspection (editing and other verification procedures), rework when inspection results so indicate, rework as a result of large errors in databases and published material, adjustments of complaints from survey buyers (nonresponse rate too high, product not delivered on time or not delivered according to specifications), production of data that are not used due to lack of relevance or other deviations from customer needs, and costs for running processes that are not sufficiently trimmed (for instance, managers often worry about redundant labor and abstain from making improvements).

Some of these failure and appraisal costs would disappear if processes were designed to generate virtually error-free data. The prevention costs can then be seen as the investment necessary to achieve this preferred state of affairs.

As pointed out in Juran and Gryna (1980) it is essential to quantify quality costs. Common cost categories used in statistics production might include planning, production, quality control, analysis, publication and dissemination. But in statistics production there are certain costs associated with making, discovering, correcting, and preventing errors or other defects. These cost categories differ from ones usually used, and securing cost figures for them seems essential for a proper allocation of improvement efforts.

8. Summary

We have tried to make a case for assuring survey quality by continuously improving survey processes so that they become stable and predictable. The ultimate goal is to have processes that are efficient and at the same time generate very small error rates. This view should be contrasted to the usual for assuring survey quality strategy involving lots of inspection and post-survey evaluation efforts.

For most organizations changing to a TQM environment with a process perspective is not easy. It is quite common that barriers exist between parts of an organization and that different subcultures exist. Many claim they "know" what the customer wants and that there is no need for special data collection. In statistical organizations a common attitude is that we have always worked "like this" and always been "for quality."

When it comes to the work itself the team approach might be an unfamiliar thing to some people. Problem-solving by experts who deliver the solutions to those who work on the processes is still very common but experience shows that such solutions are difficult to implement. Also, measurements of key process variables and the capability to distinguish between special and common causes of variation are often lacking.

Despite these roadblocks we have seen examples of how the process perspective can be successful. The recent monograph on Survey Measurement and Process Quality (Lyberg et al 1997) contains lots of examples of this alternative thinking.

9. References

Andersson, C., Engström, P., Granqvist, L., Jansson, C., and Ullberg, A. (1997). Granska effektivt. Statistics Sweden. (Efficient Editing) In Swedish.

Batcher, M. and Scheuren, F. (1997). CATI Site Management in a Survey of Service Quality. In Survey Measurement and Process Quality, Lyberg, L. et al, Chapter 25, Wiley.

Biemer, P. and Caspar, R. (1994). Continuous Quality Improvement for Survey Operations: Some General Principles and Applications. Journal of Official Statistics, 10, 307-326.

Colledge, M. and March, M. (1993). Ouality Management: Development of a Framework for a Statistical Agency. Journal of Business and Economic Statistics, 11. 157-165.

Couper, M., Holland, L., and Groves, R. (1991). Developing Systematic Procedures for Monitoring in a Centralized Telephone Facility. Journal of Official Statistics, 8, 63-76.

Deming, W.E. (1986). Out of the Crisis. Cambridge, MA: MIT Press.

Dillman, D. (1978). Mail and Telephone Surveys. Wiley.

Dodge, H.F. and Romig, H.G. (1944). Sampling Inspection Tables. Wiley.

Groves, R. and Couper, M. (1998). Nonresponse in Household Interview Surveys. Wiley.

Hansen, M., Hurwitz, W.N., and Pritzker, L. (1964). The Estimation and Interpretation of Gross Differences and the Simple Response Variance. In Contributions to Statistics, presented to professor P.C. Mahalanobis on the occasion of his 70th birthday, 111-136.

Japec, L., Ahtiainen, A., Hörngren, J., Lindén, H., Lyberg, L., and Nilsson, P. (1997). Minska bortfallet. Statistics Sweden. (Reduce the Nonresponse Rate) In Swedish.

Juran, J.M. and Gryna, Jr, F. (1980). Quality Planning and Analysis. McGraw-Hill.

Lyberg, L., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds.)(1997). Survey Measurement and Process Quality. Wiley.

Lyberg, L. (1981). Control of the Coding Operation in Statistical Investigations - Some Contributions. Series Urval, no. 13. Statistics Sweden.

Minton, G. (1972). Verification Error in Single Sampling Inspection Plans for Processing Survey Data. Journal of the American Statistical Association, 67, 337, 46-54.

Morganstein, D. and Hansen, M. (1990). Survey Operations Processes: The Key to Quality Improvement. In Data Quality Control: Theory and Pragmatics, G.E. Lupins and V.R.R. Uppuluri (eds.) New York: Marcel Dekker, 91-104.

Morganstein, D. and Marker, D. (1997). Continuous Quality Improvement in Statistical Organizations. In Survey Measurement and Process Quality, L. Lyberg et al (eds.) Wiley, 475-500.

Scholtes, P. (1994). The Team Handbook. Joiner Associates Inc.

Schwarz, N. (1997). Questionnaire design: The Rocky Road from Concepts to Answers. In Survey Measurement and Process Quality, L. Lyberg et al (eds.) Wiley, 29-45.

U.S. Bureau of the Census (1972): Evaluation and Research Program of the U.S. Census of Population and Housing, 1960: Effects of Coders, Series ER 60, no. 9. Government Printing Office.