

HOW IMPORTANT IS ACCURACY?

Fritz Scheuren¹

ABSTRACT

Accuracy is a product or service characteristic. The way we achieve accuracy and quality attributes like timeliness, relevance, accessibility, etc., speak to process quality. Deming's ideas and those of Juran and many others are touched on in this discussion to support the distinction and to disentangle these two kinds of "quality," both of which have been themes at the Conference.

KEY WORDS: Achieving Process Quality , Continuous Improvement Dimensions, Systems Thinking, Understanding Process Variability, Listening to the Voices of the Customer and of the Employee

1. PROCESS QUALITY AND ACCURACY

Accuracy, as a product or service characteristic, is a major goal of all statistical organizations. There is wide agreement that whatever else such organizations do they must meet minimal accuracy standards (e.g., Kalton, 2001). For great statistical organizations, like Statistics Canada, it would be unthinkable to operate otherwise (Brackstone 1999, 2001). There are other goals, mentioned by my fellow panelists and in other sessions at this Conference, that also have received wide acceptance as essential quality attributes – timeliness, relevance, accessibility, etc. These are just below accuracy in importance but still essential requirements that statistical organizations must strive for.

The purpose of the present paper is to look briefly at the process that gets employed to achieve these and other quality goals. "Process Quality" is one way to title what will be discussed. A barrier that needs to be overcome in this discussion is the use of the word "quality" in these two very different ways -- both as a goal and as a means. Frankly I dislike the word "quality" because of this double meaning and the even greater confusion between both of these workaday definitions of quality and the use of the word in a third emotion-laden sense. But we are stuck with the problem and so let me make a few distinctions and introduce my main points along the way.

Quality is clearly an overused word that has lost much of its motivational punch. Ironically, the hype around the word quality, especially a few years ago, may have been a major barrier to improving practice. After all, who would tell you they are not trying to do a quality job? As methodologists, we want measurable quantities. The use of the word "quality" as a "feel good" or "feel bad" will, therefore, not be developed further here.

Two serviceable definitions of quality that do connect quality means and quality ends are "fitness for use" and "conformance to requirements." Each of these links in different ways the role of the employee doing the work (conformance to requirements) to the client receiving the product or service (fitness for use). When used together these two methods of communicating quality can create efficient systems that achieve the desired accuracy level (say) or other quality attribute. Many papers at this conference employed one or the other of these definitions.

¹ Fritz Scheuren, NORC, University of Chicago, scheuren@aol.com.

My approach here will be to build on ideas inspired by Deming (1986) and Juran (1989) in discussing process quality. Specifically, following Deming, I will group my remarks under the headings: systems thinking, understanding variability, process measurement, and most importantly on the need for constancy of purpose within the organization to stay focused on continuous improvement.

2. SYSTEMS THINKING

Deming's notion of "system" can go a long way towards explaining how inclusive the idea of process quality is. Obviously data producers inside statistical organizations are all part of society's knowledge-creation system. Notice I do not characterize "the system" as producing data but knowledge, although data can be one of the system's intermediate or even final products. Clients share the system with producers. They are not outside of it. They need to be integral to it; otherwise how can the fitness for use definition of quality be actionable? It is, in fact, systems thinking that links the two definitions of process quality: the employee doing the work (conformance to requirements) and the client receiving the product or service (fitness for use).

We are providing a service and not a product (e.g., Fellegi 1999); thus, simply using data producer measures of product quality will not be enough, no matter how good they are. In assessing the quality of a service, there are three elements to look at, which can be expressed in equation form as:

$$\text{Quality} = \text{Producer} + \text{Client} + \text{Joint.}$$

In most of the statistical literature on quality, the focus has been on the first term. Our (appropriate) attention at this session to accuracy would be an example. What we need to emphasize more, however, is the third (interaction) term. It is the key to systems thinking (Batcher and Scheuren 1997).

Two-way client-employee communication and real listening are entirely too rare. Right now, much of the producer-client communication that goes on is systematically carried out in only one direction. Producers need to better learn the language of their customers and especially to more frequently employ listening mechanisms, like customer surveys or focus groups, that get needed feedback. Failure to listen to all parts of the system may be among the most serious barriers to improving process quality. The idea that the creation of information or knowledge is done jointly with clients, despite being a point accepted by all, has not penetrated deeply enough into many producer operations.

Many statistical organizations do an extremely good job of providing detailed information about their processes and running training classes for users. All kinds of interesting general tidbits can be had, but not always what they might have meant to the interpretation of a particular client result. The need for a "just-in-time" knowledge, like that possible with web-based metadata systems (e.g., Dippo and Sungren 2000), can be a partial answer. I was glad to see how much progress has been made in this area, as evidenced by the several papers given on this topic over the last few days. Without a doubt, a holistic approach is needed and appears to be coming, similar in focus to the goal of a quality profile report (e.g., Jabine 1994, Kalton et al. 2000).

Fair enough, we need to systematically listen to what has been overdramatized as the "Voice of the Customer" and the "Voice of the Employee." But what is the topology of this listening? It is common to characterize a traditional statistical system as beginning with what a client wants. Then a handoff occurs to various producer components (samplers, questionnaire designers, field staff, etc.) who one after another in a roughly sequential fashion, "do their thing" until the end result is provided back to the client. This can be characterized as a system having a *chain link* topology.

In a chain link system, the listening to the client occurs basically at the beginning and the end. One immediate problem is that most of the employees working within the system never directly hear the “Voice of the Customer” at all. And what about the “Voice of the Employee”? Who hears this “Voice”? Even if you have “driven out fear” the employee may be listening only within their own link in the chain -- their own department. To have sustainable process quality you also have to follow another of Deming’s 14 points. In addition to “driving out fear” you have to “break down barriers between departments.”

The system’s topology that best fits this world is that of a *suspension bridge*, with the driving surface (work load) supported by cables (work processes) hung from tall towers (organizational goals) firmly rooted in bed rock (core societal values). The cables are the former chain links stretched out and touching at every point, so every employee is directly in touch with every other employee and with their clients, both internal and external. To my knowledge this topology has yet to be employed in large statistical agencies, although elements exist now given the web access that customers and employees can have to each other. Because of the changing technologies we use the opportunity to get closer to this structure can present itself, however, and should be grasped.

3. UNDERSTANDING VARIABILITY

As statisticians we can rightly claim that we understand variability, certainly not all about it but a great deal. In our data collection we go to great efforts to find ways to measure and reduce it. Maybe in this area Deming has nothing to say to us that we do not already know? Frankly in large part I would agree. Statistical agencies obviously do understand variability. It is a core strength. But maybe they have the “cobbler’s children” problem? Perhaps they could apply what they know more to their own internal problems?

Well my answer to this is “Yes”-- more could be done to use our tools on ourselves and not just on our data. But much is already going on (whether in government or not). This very Conference would be an example of how we are learning from variability, variability of approach and of result. Opportunities to benchmark processes, one organization against another, are now being taken more and more. The recent Quality Conference sponsored by Statistics Sweden in May of this year showed an impressive amount of cooperation and cross-organizational learning. I would like to see more benchmarking, though, between governmental statistical organizations and private sector ones. Especially firms who, incidentally or as their main mission, are also in the knowledge-creation business. Graham’s and my presence on the program today speaks to this directly, of course.

Let me offer one worry. I have said that my assessment is basically positive. I do like the use now being made by government statistical organizations of their own internal variability and their benchmarking against alternative approaches. Both of these can help them achieve greater accuracy at a lower cost. I do wonder, though, if the speed of innovation is fast enough. Government organizations suffer greatly from an apparent lack of competition. Budget problems can and are used to avoid changes that might lead to improvements (I have been guilty of this myself when I was in government).

Often in such a world, the case to rethink a process that seems to have been fairly successful historically may not be seen as compelling. Clients may not even realize that they are not fully utilizing statistical results. They may ignore nonsampling errors altogether, for example, or treat them just verbally, without any quantification or sensitivity analyses. While I have stressed above the need to include the client in the system, many clients simply do not want this level of involvement.

I used the word “apparent” above in referring to the lack of competition. Why? My reason for this word is that there is competition for employees and within the society as a whole for resources. Let me elaborate only slightly, since many of you here are well aware of the issues.

First, employees are increasingly highly educated and skilled and will be sought after by the private sector. What will attract or keep them in government? One answer is more empowerment to change work processes. This lets employee variability come out as a strength. How do we respect their individual variability and use it as a strength? How or rather where should we try to reduce it? Creating strongly shared work values is part of the answer. Employee work values is where variability must be reduced. In other words, using values as a control and de-emphasizing bureaucratic checks is the way to go. This is where that wonderful Japanese work, KAISEN, comes in -- representing, as it does, that a culture of continuous improvement be fostered and internalized in each employee.

Second, the market share of the knowledge business that governments have in western style democracies is falling dramatically. I see no end in sight. Substitutes are being found for traditional government products which often come to market too late to be used as more than checks on other more available sources. What should the response be? And how does this connect to variability? Of course improving existing delivery systems by improving process quality is part of the answer. But a wider array of government efforts could be part of the response too -- hence an increase in the variability of what is done. Being more outward looking is another aspect. The sponsorship of great journals like *Survey Methodology* and the *Journal of Official Statistics* would be one example of a successful outward looking approach. Linking continuing government niches to wider society knowledge generation efforts is obvious too. That is, create hybrid products or services that combine internally generated information with that available outside government.

4. PROCESS MEASUREMENT

Process quality, as we have been using it, has an action orientation that it shares with the word KAISEN. Measuring process quality must, therefore, focus in large part on improving systems. What are the biggest barriers to improving process quality measurement? Ironically we may need to make quality measures more long-run or systems-oriented and less focused on immediate specific product or service results. This is seemingly a violation of the conditionality principle but no more, I would argue, than is a confidence interval statement.

Most of the issues here have been well worked by others (e.g., Platek and Särndal 2001; U.S. Federal Committee on Statistical Methodology 2001), so I will be brief. Let me just emphasize a few of the barriers to making quality measurement more systems-oriented:

(1) In an information age, producers alone cannot judge fitness for use; hence, the system of assessment must be open to clients. Producers do not usually make it possible for clients to do their own assessments of nonsampling error. Because of confidentiality concerns, some survey public use files do not even make it possible to directly estimate sampling error. What clients really want to know is how to deal with defects remaining after all survey processing, and this is often the one thing that our current systems cannot readily help them learn.

(2) Existing process quality measures are often raw by-products of survey operations and do not use either modern computational power or, indeed, to any serious degree, the power of modern statistics. In fact, simply reciting a long list of places where failures to conform to requirements occurred can create a misleadingly bad impression of quality. A major effort is made in most government statistical work to eliminate inconsistencies, impute missing items, and adjust for nonresponse and coverage errors. None of these efforts get "scored" in the usual quality measures.

(3) Major language barriers exist between clients and employees. Some of these arise because of producer emphasis on randomization-based inference (Hansen, Tepping and Madow 1988) and a consequent avoidance of models. Other issues are more pervasive and go to a lack of enough statistical thinking among both clients and employees. This lack and the language barrier it creates

may be the toughest problem to be faced. Listening does not lead to understanding without a common language.

Juran (1988) has emphasized that there really are two parallel systems, the actual operating system and the quality measurement system that sits alongside it. More attention must be paid to this second system, as a system or (to continue our metaphor) as a strand in the cable carrying the work load. Frustration with current process quality measures may grow out of the department-by-department nature many of them have. Clients need to know when the system may be said to have “failed safe.” An overall, in this case, Juran-inspired systems perspective can help here, albeit it is hard to operationalize (but see Scheuren 2000). The improvements in web-based access should help, provided the existing language barriers are worked on.

The language and related statistical thinking issues raised above are too big for a full discussion. Certainly I have no “quick fix.” But let me say something, as a methodologist, about models. Randomization-based thinking is insufficient alone. It has a role but a wider basis for practice is needed, as we understand nonsampling issues more and more. The use of explicit informative models may, in and of itself, help clients, who often operate in a model-dependent world anyway. And with producer help, clients may be able to better integrate producer knowledge, as embedded in models, with subject matter knowledge also expressed in models.

5. CONSTANCY OF PURPOSE

Deming, as the story goes, was asked once which of his 14 points was the most important. He thought for a while. So long, in fact, that people began to offer suggestions. Finally, he settled on “Constancy of Purpose.” If you want to achieve quality you must be persistent and never ceasing in giving attention to process improvement. Crises are ubiquitous, especially early on. Attention to process quality can be easily lost and often is.

Many organizations and individuals have started making process quality improvements; especially after a failure. Some are still on-path but efforts like TQM are largely a thing of the past. On the other hand I heard a great many speakers talking about process improvements that might have been labeled TQM a few years ago. Deming would not have liked TQM, by the way, he hated phoniness and hype and TQM had plenty of both.

So how can we know that the emphasis I have given to this topic is practical and worth looking into more. Why do I think I’m right? My answer would be to look around you. Many of the people in this room have given a lifetime to our profession or plan to. A few have already showed that constancy of purpose can lead to greatness. Jon Rao would be one example of a person who lives it and leads others to follow. Thank you Jon. Let me also thank the speakers at this Conference for renewing my own commitment, wherever that leads.

REFERENCES

Batcher, M., and F. Scheuren. 1997. “CATI Site Management in a Survey of Service Quality,” *Survey Measurement and Process Quality*. Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, Trewin, eds. John Wiley & Sons, Inc.: New York, 573–588.

Brackstone, G. (1999). “Managing Data Quality in a Statistical Agency,” *Survey Methodology*, 25, pp. 139-149.

- Brackstone, G. (2001). "Panel Discussion - How Important is Accuracy?" *Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: a Methodological Perspective*.
- Deming, W. E. (1986). *Out of the Crisis*, MIT Press.
- Dippo, C., and B. Sundgren. 2000. "The Role of Metadata in Statistics." *See also* Colledge, M., and E. Boyko. 2000. "Collection and Classification of Statistical Metadata: The Real World of Implementation." Both of these papers were presented at the Second International Conference on Establishment Surveys in Buffalo.
- Fellegi, I. P. 1999. "Statistical Services: Preparing for the Future," *Survey Methodology* 25 (2): 113–128.
- Jabine, T. 1994. *Quality Profile for SASS: Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)*. National Center for Education Statistics.
- Juran, J. M. 1988. *Juran on Planning for Quality*. New York: Free Press.
- Kalton, G., M. Winglee, S. Krawchuk, and D. Levine. 2000. *Quality Profile for SASS Rounds 1–3: 1987–1995, Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)*. National Center for Education Statistics.
- Kalton, G. (2001). "How Important Is Accuracy?" *Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: a Methodological Perspective*.
- Platek, R. and Särndal, C.-E. (2001), "Can a Statistician Deliver?", *Journal of Official Statistics*. See also the excellent discussion.
- Scheuren, F. (2000). "Quality Improvement through Improved Quality Measurement" *Proceedings of the American Statistical Association, Section on Survey Research Methods*.
- U.S. Federal Committee on Statistical Methodology, Subcommittee on Measuring and Reporting the Quality of Survey Data (2001), *Measuring and Reporting Sources of Error in Surveys*, Statistical Policy Working Paper 31, Washington, D.C.: Statistical Policy Office, Office of Management and Budget.