Journal of Official Statistics
Vol. 8. No. 2, 1992, pp. 119–128

Research and Training for Quality Improvement[1]

George E.P. Box[2]

Abstract: Quality improvement is seen as an extension and democratization of use of the scientific method. The domain of such an extension has a number of dimensions that include users, disciplines, areas of endeavor, time (never-ending improvement) and the evolving discovery of new factors affecting the system. Simple tools are described, which may be used by the whole work force for studying the data continuously generated by the process, to achieve improvement. More sophisticated methods, particularly the basics of experimental design, may be used by the engineer or scientist to greatly increase the efficiency of experimentation whether intended to increase the mean, reduce the variation, or to discover designs for process and product which are robust to both component variation and environmental change.

Key words: Quality improvement; scientific method; creativity; design of experiments.

1. Introduction

Quality improvement is about finding out how to do things better. The efficient way to do this is by using scientific method – by which I mean a group of techniques which catalyze the acquisition of knowledge – often employed in the past by only a small elite of trained researchers. Modern quality improvement extends and democratizes the domain of scientific method over a number of dimensions (Box 1989b) which include:

users:	from the chief executive officer to the janitor;
disciplines:	e.g., management, statistics, engineering, psychology, anthropology;
areas of endeavor:	e.g., factories, hospitals, airlines, department stores;
time:	never-ending quality improvement;
causative factors:	an evolving panorama of factors affecting the operation of the system often not known in advance but needing to be *discovered*.

1.1. Users

In the past, we have rationed scientific method and provided a license to use it only to those having research degrees. But in fact,

[1] This paper is based on a presentation at the American Statistical Association's Joint Statistical Meetings, August 6–10, 1989, Washington, DC, U.S.A.

[2] University of Wisconsin-Madison, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, WI 53705, U.S.A.

Acknowledgement: This research was sponsored by the National Science Foundation under Grant No. DMS-8420968, the United States Army under Contract No. DAAL03-87-K-0050, and by the Vilas Trust of the University of Wisconsin-Madison, U.S.A.

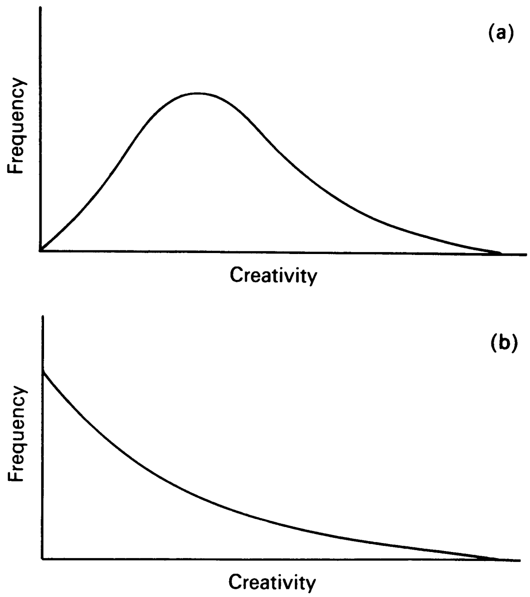

Fig. 1. (a) Frequency distribution of workers by creativity. (b) Frequency distribution of problems by needed creativity.

the single entity which divides the human race from the rest of the animal kingdom is its creativity and just as fish must swim and birds must fly, it is natural for human beings to be creative.

To understand the implications of this (Box 1989a), imagine two frequency distributions. Figure 1a shows the kind of frequency distribution of creativity we might expect to find among the workforce of some organization such as a factory, a hospital, a university, or an airline. This is the kind of distribution usually found for human skills, showing clustering about a modal value with falling off in frequency on either side. Traditionally it was assumed that the highly qualified few fell in the right tail of this distribution and these few were therefore the "licensed" problem solvers.

Figure 1b shows a frequency distribution of problems that might beset an organization, classified by the degree of creativity needed to solve such problems. Here one expects a Pareto-like distribution with a large number of very simple problems and the frequency falling off as the problems become more challenging.

In the past, the designated problem solvers were too few to deal with any but a small proportion of the problems and were in any case over-qualified to deal with most of them. But, one ought not need a Ph.D. to make sure, for example, that hospital records are available when a patient comes to be examined. Most of the problems that cause inefficiency in organizations are simple problems and the vast pool of creativity potentially available to solve them is not used.

1.2. Disciplines

Quality does not fit into any one slot. It is not just statistics, not just engineering, not just management, not just psychology. Thus we need to think always in terms of cooperation, not departmentalism, and of teams, not of lone rangers. We must listen carefully, because frequently what we have to absorb is from a field other than our own. Also while retaining our humility, we must, at the same time, not be overawed.

1.3. Areas of endeavor

Quality improvement is most often thought of as applying to manufacture on the factory floor. But even in manufacturing organizations a high proportion of the workforce are otherwise engaged – in billing, invoicing, planning, scheduling, for example; all of these operations can be improved. But outside such industrial organizations, all individual citizens must deal with a complex world involving hospitals, government departments, universities, airlines, and so forth. Lack of quality in these organizations causes needless expense, wasted time, and unnecessary frustration. Quality improvement applied to these activities could free us all for more productive and pleasurable pursuits.

1.4. Time

For never-ending improvement there must be a long-term commitment to renewal. A commonly used statistical model links a set of variables "known" x_k with a response y by an equation $y = f(x_k) + e$ where e is an error term, often imbued by statisticians with properties of randomness, independence and normality. A more realistic version of this model is

$$y = f(x_k) + e(x_u)$$

where x_u is a set of variables whose nature and behavior is unknown. By skillful use of the techniques of informed observation and experimental design, as time elapses, elements of x_u are transferred into x_k – from the unknown into the known. This transference is the essence of modern quality improvement and has two consequences:

a. once a previously unknown variable has been identified it can be fixed at a level that produces the best results;
b. by fixing it we remove an element which previously contributed to variation.

1.5. Factors and assignable causes

The field of factors potentially important to quality improvement also can undergo seemingly endless expansion. Problems of mathematical optimization are frequently posed as if they consisted of maximizing some response y over a *known* k-dimensional space of factors x_k, but in practice the factor space is never totally known and is continually *developing*.

2. Requirements for Change

This extension of the dimensions and the domain for the application of scientific method represents a very great change from the way that things are usually done. It requires (see, for example, Deming 1986) the following:

a. change in management attitudes and structure to produce and preserve the designed change in the organization;
b. management training to achieve this;
c. concomitant change in each employee's job;
d. training to equip each employee to carry out his/her changed role.

Considerations (a), (b), (c), and (d) are all of great importance. Nothing can happen without getting, not only the support of management, but its enthusiastic *involvement*. But it is equally true that even if management is genuinely ready and equipped to move, it will not get anywhere unless the necessary changes in employees' roles and the necessary new training can be put in place as in (c) and (d) above. I will focus on these two considerations.

2.1. Informed observation and design experiment

Until comparatively recently technical advance was slow – the ships of the thirteenth century were somewhat better designed than those of the twelfth century but the differences were not very dramatic. And then three or four hundred years ago a process of quickened technical change began which has ever since been accelerating. This acceleration is attributed to an improved process for finding things out we call the *scientific method*.

We can, I think, explain at least some aspects of this scientific revolution by considering a particular instance of discovery. We are told that in the late seventeenth century it was a monk from the Abbey of Hautvillers who first observed that a second fermentation in wine could be induced which produced a new and different sparkling liquid, delightful to the taste, which we now call champagne. Now the culture of wine itself was known from the earliest records of man and the conditions necessary to induce the production of champagne must have occurred accidentally countless times throughout antiquity. However, it was not until this comparatively recent date that the actual discovery was made. This is less surprising if we consider that to induce an advance of this kind two circumstances must coincide. First an *informative event* must occur and second a *perceptive observer* must be present to see it and learn from it.

Now most events that occur in our daily routine correspond more or less with what we expect. Only occasionally does something occur which is potentially informative. Also many observers, whether through lack of essential background or from lack of curiosity or motivation, do not fill the role of a perceptive observer. Thus the slowness in antiquity of the process of discovery can be explained by the extreme rarity of the chance *coincidence* of two circumstances *each* of which is itself rare. It is then easily seen that discovery may be accelerated by two processes which I will call informed observation and directed experimentation.

By a process of *informed observation* we arrange things so that, when a rare potentially informative event *does* occur, people with necessary technical background and motivation are there to observe it. Thus, when recently a supernova exploded, the scientific organization of this planet was such that astronomers observed it and learned from it. A quality control chart fills a similar role. When such a chart is properly maintained and displayed, it ensures that any abnormality in the routine operation of a process is likely to be observed and associated with what Shewhart called an *assignable cause* – so leading to the gradual elimination of disturbing factors.

A second way in which the rate of acquisition of knowledge may be increased is by what I will call *directed experimentation.* This is an attempt to *artificially* induce the occurrence of an informative event. Thus, Benjamin Franklin's plan to determine the possible connection of lightning and electricity by flying a kite in a thunder cloud and testing the emanations flowing down the string, was an invitation for such an informative event to occur.

Recognition of the enormous power of these methods of scientific advance is now commonplace. The challenge of the modern movement of quality improvement is nothing less than to use them to further, in the widest possible manner, the effectiveness of human activity.

2.2. *Informed observation*

The less sophisticated problems in quality improvement can often be solved by informed observation using some very simple tools that are easily taught to the workforce. While on the one hand, Murphy's law implacably ensures that anything that can go wrong with a process will eventually go wrong, the same law also ensures that every process produces information that can be used for its own improvement. In this sentence the word "process" could mean an industrial manufacturing process, or a process for ordering supplies or for paying bills. It could also mean the process of admission to a hospital or of registering at a hotel or of booking an airline flight.

Three simple strategies that can be used to defeat Murphy are *Corrective Feedback*, *Preemptive Feedforward*, and *Simplification* (Box 1989b).

Corrective Feedback occurs when the study of system faults leads to their eradication. Preemptive Feedforward occurs when careful forethought prevents faults from occurring. Simplification occurs when unnecessary complication in a process is unravelled (see, for example, Fuller 1985).

To put such strategies into effect, the workforce needs a simple set of tools such as is described by Ishikawa (1976) in his invaluable little book available in English and written for foremen and workers to study together. I would make a few new additions, so that my list would read:

1. Check sheets and recordkeeping
2. Pareto charts
3. Flow diagrams
4. Cause–effect diagrams
5. Run charts
6. Histograms
7. Stratification
8. Scatter plots
9. Cusums
10. Other graphs.

Cusums are here intended to be used simply as a diagnostic graphical tool pointing to approximate times at which changes in mean (or in range) may have occurred, and so helping with the assignment of cause. No formal assessment of significance would be needed at this stage. Various illustrations of the use of these tools will be found, for example, in the Ishikawa and Fuller references above and also in Box and Bisgaard (1987).

2.3 Experimental design

The solution by the workforce of a multitude of simple problems using the elementary tools listed above can produce major improvements in quality and productivity, but there is still another area where additional major gains can be accomplished.

Although R.A. Fisher invented experimental design more than 60 years ago, its use has still not become customary in industrial experimentation. This is partly the fault of the universities. The use of one factor at a time experimentation to study multifactor situations has (with rare exceptions) become indefensible, yet it is still taught in departments of chemistry and engineering and is still prevalent in industry.

The excuse usually advanced by engineering colleges for not teaching experimental design is that the syllabus is already full and that to include a course on design and analysis of experiments would mean that something would have to be taken out. "What could we take out?" I am often asked. To this I reply: "Take out almost anything; surely nothing is more important than to equip engineers to properly conduct investigations and to run experiments." As I was told by a senior manager in a highly successful Japanese company, "an engineer not trained in the design of experiments is not an engineer."

3. Which? How? Why?

Suppose y is some quality characteristic whose probability distribution depends on the levels of a number of factors x. Experimental design may be used to reveal certain aspects of this dependence; in particular how the mean $E(y) = f(x)$ and the variance $\sigma^2(y) = F(x)$, depend on x. Both choice of design and method of analysis are greatly affected by what we know or what we think we know about the input variables x and the functions $f(x)$ and $F(x)$ (see, for example, Box, Hunter, and Hunter 1978).

Which: In the early stages of investigation the task may be to determine *which* subset of variables x_k chosen from the larger set x

are of importance in affecting y. In this connection a Pareto hypothesis (a hypothesis of "effect sparsity") becomes appropriate and the projective properties into lower dimensions in the factor space of highly fractionated designs (Box, Hunter, and Hunter 1978; Finney 1945; Plackett and Burman 1946; Rao 1947) may be used to find an active subset of k active factors. Analyses based on normal plots (Daniel 1959) or Bayesian methods (Box and Meyer 1986a) which take account of the Pareto hypothesis are efficient and geometrically appealing.

How: When we know, or think we know, which are the important variables x_k we may need to determine more precisely how changes in their levels affect y. Often the nature of the functions $f(x_k)$ and $F(x_k)$ will be unknown. However, over some limited region of interest, a local Taylor's series approximation of first or second order in x_k may provide an adequate approximation; particularly if y and x_k are re-expressed when necessary in appropriate alternative metrics. Fractional factorials and other response surface designs of first and second order are appropriate here. Maxima may be found and exploited using steepest ascent methods followed by canonical analysis of a fitted second degree equation in appropriately transformed metrics (Box and Wilson 1951). The possibilities for exploiting multidimensional ridges and hence determining alternative optimal processes become particularly important at this stage (see, for example, Box and Draper 1986).

Why: Instances occur when a mechanistic model can be postulated. This might take the form of a set of differential equations believed to describe the underlying physics. Various kinds of problems arise. Among these are:

How should parameters (often corresponding to unknown physical constants) be estimated from data?
How should candidate models be tested?
How should we select a model from competing candidates?
What kinds of experimental designs are appropriate?

Workers in quality improvement have so far been chiefly occupied with problems of the "which" and occasionally of the "how" kind and have consequently made most use of fractional factorial designs and other orthogonal arrays, and of response surface designs.

4. Studying Location, Dispersion, and Robustness

Experimental design has been used most often as a means of discovering how the process might be changed to increase the *mean* of some quality characteristic. It has also an important role to play in reducing dispersion.

4.1. *Using experimental designs to minimize variation*

High quality, particularly in the parts industries (e.g., automobiles, electronics), is frequently associated with minimizing dispersion. In particular, the simultaneous study of the effect of the variables x on the variance as well as the mean is important in the problem of bringing a process on target with smallest possible dispersion (Phadke 1982).

Bartlett and Kendall (1946) pointed to the advantages of analysis in terms of log s_y^2 to produce constant variance and increased additivity in the dispersion measure. It is also very important in such studies to remove transformable dependence between the mean and standard deviation. Taguchi (1986, 1987) attempts to do this by the use of signal to noise ratios. However, it may be shown that it is much less restrictive, simpler, and more statistically efficient to proceed by direct data transformation obtained, for example, by a lambda plot (Box 1988).

A practical difficulty may be the very large number of experimental runs needed in such studies if complicated designs are employed. It was recently shown how, using what Fisher called hidden replication, unreplicated fractions may sometimes be employed to identify sparse dispersion effects in the presence of sparse location effects (Box and Meyer 1986b).

4.2. *Experimental design and robustness to the environment*

A well designed car will start over a wide range of conditions of ambient temperature and humidity. The design of the starting mechanism may then be said to be "robust" to changes in these environmental variables. Suppose $E(y)$ and possibly also $\sigma^2(y)$ are functions of certain design variables x_d which determine the design of the system and also of some environmental variables x_v which, except in the experimental environment, are not under our control. The problem of robust design is to choose a desirable combination of design variables x_{do} at which good performance is experienced over a wide range of environmental conditions.

Similar problems were earlier discussed by Youden (1961a, b), Michaels (1964), and Wernimont (1977) and recently their importance in quality improvement has been emphasized by Taguchi. His solution employs an experimental design which combines multiplicatively, an "inner" design array and an "outer" environmental array. Each piece of this combination is usually a fractional factorial design or some other orthogonal array. Recent research has concentrated on various means for reducing the burdensome experimental effort which currently may be needed for studies of this kind.

4.3. *Robustness of an assembly to variation in its components*

In the design of an assembly, for example an electrical circuit, the exact mathematical relation $y = f(x)$ between the quality characteristic of the assembly, such as the output voltage y of the circuit and the characteristics x of its components (resistors, capacitors, etc.), may be known from physics. However, there may be an infinite variety of configurations of x that can give the same desired mean level $E(y) = \eta$, say. Thus an opportunity exists for optimal design by choosing a "best" configuration.

Suppose the characteristics x of the components vary about "nominal values" ξ with *known* covariance matrix $\mathbf{V}$. Thus, for example, a particular resistance x_i might vary about its nominal value ξ_i with known variance σ_i^2. (Also variation in one component would usually be independent of that of another so that $\mathbf{V}$ would usually be diagonal.) Now variation in the input characteristics x will transmit variation to the quality characteristic y so that for each choice of component nominal values ξ which yield the desired output $y = \eta$ there will be an associated mean square error $E(y - \eta)^2 = M(\eta) = F(\xi)$.

Using a Wheatstone Bridge circuit for illustration, Taguchi and Wu (1985) pose the problem of choosing ξ so the $M(\eta)$ is minimized. To solve it they again employ an experimental strategy using inner and outer arrays. Box and Fung (1986), see also Morrison (1957), have pointed out, however, that their procedure does not in general lead to an optimal solution and that it is better to use a simpler and more general method employing a standard numerical nonlinear optimization routine. The latter authors also make the following further points:

a. For an electrical circuit it is reasonable to assume that the relation $y = f(x)$ is known. But when, as is usually the case, $y = f(x)$ must be estimated experimentally, the problems are much more complicated and require further study.
b. It is also supposed that all of the σ_i^2 are known and furthermore that they *remain fixed or change in a known way* (for example, proportionally) when the ξ_i change. The nature of the optimal solution can be vastly different depending on the validity of such assumptions.

Taguchi's quality engineering ideas are clearly important and present a great opportunity for development. It appears, however, (see, for example, Box, Bisgaard, and Fung 1988) that the accompanying statistical methods that Taguchi recommends employing "accumulation analysis," "signal to noise ratios," and "minute analysis" are often defective, inefficient, and unnecessarily complicated. Furthermore, Taguchi's philosophy seems at times to imply a substitution of statistics for engineering rather than the use of statistics as a catalyst to engineering (Box 1988). Because such deficiencies can be easily corrected it is particularly unfortunate that, in the United States at least, engineers are often taught these ideas by instructors who stress that no deviation from Taguchi's exact recipe is permissible.

4.4. Training

Instituting the necessary training for quality is a huge and complex task. Some assessment must be made of the training needs for the workforce, for engineers, technologists, and scientists, and for managers at various levels. We must consider how such training programs can be organized using the structure that we have within industry, service organizations, technical colleges, and universities. A maximum multiplication effect will be achieved by a scheme in which the scarce talent that is available is employed to teach the teachers within industry and elsewhere.

5. Conclusions

We draw the following conclusions:

1. By organizing the workforce in problem-solving teams trained in simple graphical data analysis we can utilize, instead of throwing away, an enormous resource for creativity.
2. By equipping engineers with statistical design and graphical analysis tools accompanied by supporting computer software, we can greatly catalyze the design of new processes and improvement of old ones. Formidable management problems must be overcome to allow this to happen.

6. References

Bartlett, M.S. and Kendall, D.G. (1946). The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation. Journal of the Royal Statistical Society, ser. B, 8, 128–150.

Box, G.E.P. (1988). Signal-to-Noise Ratios, Performance Criteria and Transformations. Technometrics, 30, 1–17.

Box, G.E.P. (1989a). Quality Improvement: An Expanding Domain for the Application of Scientific Method. Philosophical Transactions of the Royal Society, ser. A, 327, 617–630. Also Report No. 31, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Box, G.E.P. (1989b). When Murphy Speaks – Listen. Quality Progress, 22, 79–84. Also

Report No. 34, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Box, G.E.P. and Bisgaard, S. (1987). The Scientific Context of Quality Improvement. Quality Progress, 20, 54–61.

Box, G.E.P., Bisgaard, S., and Fung, C.A. (1988). An Explanation and Critique of Taguchi's Contributions to Quality Engineering. Quality and Reliability Engineering International, 4, 123–131. Also Report No. 28, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Box, G.E.P. and Draper, N.R. (1986). Empirical Model-Building and Response Surfaces. New York: John Wiley.

Box. G.E.P. and Fung, C.A. (1986). Studies in Quality Improvement: Minimizing Transmitted Variation by Parameter Design. Report No. 8, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. New York: John Wiley.

Box, G.E.P. and Meyer, R.D. (1986a). An Analysis for Unreplicated Fractional Factorials. Technometrics, 28, 11–18.

Box, G.E.P. and Meyer, R.D. (1986b). Dispersion Effects from Fractional Designs. Technometrics, 28, 19–27.

Box, G.E.P. and Wilson, K.B. (1951). On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society, ser. B, 8, 1–45.

Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments. Technometrics, 1, 311–341.

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

Finney, D.J. (1945). The Fractional Replication of Factorial Arrangements. Annals of Eugenics, 12, 291–301.

Fisher, R.A. (1935). The Design of Experiments. London: Oliver and Boyd.

Fuller, F.T. (1985). Eliminating Complexity from Work: Improving Productivity by Enhancing Quality. National Productivity Review. Also Report No. 17, Center for Quality and Productivity Improvement, University of Wisconsin-Madison.

Ishikawa, K. (1976). Guide to Quality Control. Asian Productivity Organization.

Michaels, S.E. (1964). The Usefulness of Experimental Design. Applied Statistics, 13, 221–228. With discussion, 228–235.

Morrison, S.J. (1957). The Study of Variability in Engineering Design. Applied Statistics 6, 133–138.

Phadke, M.S. (1982). Quality Engineering Using Design of Experiments. Proceedings of the Section of Statistical Education, American Statistical Association, 11–20.

Plackett, R.L. and Burman, J.P. (1946). Design of Optimal Multifactorial Experiments. Biometrika, 23, 305–325.

Rao, C.R. (1947). Factorial Experiments Derivable from Combinatorial Arrangements of Arrays. The Journal of the Royal Statistical Society, ser. B, 9, 128–140.

Taguchi, G. (1986). Introduction to Quality Engineering: Designing Quality into Products and Processes. White Plains, N.Y.: Kraus International Publications.

Taguchi, G. (1987). System of Experimental Design, Vols. 1 and 2. White Plains, NY: Kraus International Publications.

Taguchi, G. and Wu, Y. (1985). Introduction to Off-Line Quality Control. Nagaya, Japan: Central Japan Quality Control Association.

Wernimont, G. (1977). Ruggedness Evaluation of Test Procedures. Standardization News, 5, 13–16.

Youden, W.J. (1961a). Experimental Design and ASTM Committees. Material Research and Standards, 1, 862–867. Reprinted in Precision Measurement and Calibration, Special Publication 300, National Bureau of Standards, Vol. 1, 1969, Ed. H.H. Ku.

Youden, W.J. (1961b). Physical Measurement and Experimental Design. Collogues Internationaux du Centre National de la Recherche Scientifique No 110, le Plan d'Experiences 1961, 115–128. Reprinted in Precision Measurement and Calibration, Special Publication 300, National Bureau of Standards, Vol. 1, 1969, Ed. H.H. Ku.

Received October 1989
Revised December 1991