UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Quality improvement from the viewpoint of statistical method

de Mast, J.

Publication date
2002

Link to publication

Citation for published version (APA):
de Mast, J. (2002). *Quality improvement from the viewpoint of statistical method*.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:11 Nov 2022

3 Corroboration of the proposed framework

The proposed methodological framework for statistical improvement strategies, as it is described in the previous chapter, is summarised in table 2.1 (concepts **C1** through **C3**, heuristics **H1** through **H5**, and methodological rules **R1** through **R4**). Before it can be accepted, however, it should prove its value in a confrontation with the praxis of quality improvement.
The acceptability of the proposed framework has the following two aspects:

1. Is the approach that the framework generates for a given improvement project 'correct'?
2. Can the framework reconstruct approaches for improvement projects which are known to be 'correct'?

Different from empirical inquiry, there is in reconstruction research not some sort of 'objective' reality to which 'correctness' could be tested. Instead, the researcher depends on the judgement of the users and experts of the system of rules that is studied. Ideally, the proposed framework would be tested by having a number of inquirers follow the rules and heuristics in the framework to conduct their improvement projects (to test 1.) and by reconstructing a large number of projects whose approach is accepted to be 'correct' (to test 2.). Unfortunately, the first proposal is not feasible for practical reasons. For the second I could use case studies in literature, but these tend not to be presented in a raw and detailed format, but in an interpreted and streamlined form. Instead, I present the following material to corroborate the proposed framework:

- *Application to current improvement strategies*: I study to what extent the proposed framework provides an adequate reconstruction of well-known improvement strategies (to test 2.).
- *Two case studies*: As a statistical consultant I have conducted and supervised a number of improvement projects. I conducted two projects according to earlier versions of the framework[1]. The case studies of these projects present the resulting approach, the adequacy of which is open for criticism from researchers and quality professionals (thus testing 1.).
- *Further experiences in consultancy work*: During the research underlying chapter 2 provisional versions of the methodological framework were confronted over and

[1]More specific, these projects follow the approach that is described in Does, Van den Heuvel, De Mast, Schippers, Trip and Wieringa (2001).

over again with projects that I had supervised. Refraining from listing all of this material, I present some highlights which I consider illustrative for the success or the complications encountered while reconstructing these projects according to the proposed framework.

3.1 Application to current improvement strategies

In the current section I confront the proposed framework with a number of well-known statistical improvement strategies. The purpose of this collation is:

- To study to what extent a number of well-known statistical improvement strategies conform with the elements of the proposed framework. Where strategies deviate from the proposed framework, these deviations are discussed and evaluated.
- To show that the proposed framework facilitates an intelligent discussion of the methodological aspects of improvement strategies.

The material in this section is largely based on De Mast, Schippers, Does and Van den Heuvel (2000). In this article the authors compare four improvement strategies: Statistical Process Control (SPC) (Wheeler and Chambers, 1992), Taguchi's methods, the Shainin System and the Six Sigma programme. SPC, which is a strategy for quality control rather than quality improvement, is left out of the current discussion. The criteria for the selection of these strategies are: that they comply (to a considerable extent) with the definition of statistical improvement strategies given in chapter 1; that they have proven to be successful; and that they are well-known.

3.1.1 The improvement strategies

The Shainin System

Dorian Shainin put several techniques — both known and newly invented — in a coherent stepwise strategy for problem solving in a manufacturing environment. This strategy is called the Shainin System. The system is described in various papers (P. Shainin, 1993; R. Shainin, 1993). Part of the strategy is promoted by Bhote (1991). Both Shainin, but especially Bhote present the Shainin System as an alternative for SPC and Taguchi's methods. Since elements of the Shainin System are legally protected as Service Marks and some methods are rarely discussed in literature, it is difficult to obtain a complete overview.

Starting from a problem in the output of a process, the objective of the strategy is to select the one, two or three dominant causes of variation (called the Red X, Pink X and Pale Pink X, respectively) from all possible causes (the X-es). This is achieved by a 'homing in' method: using statistical analysis tools, the classes of causes in which the important causes are likely to be found are selected, thus zooming in on the Red X. Once the Red X is identified, either an irreversible corrective action is taken, or the tolerances on the Red X are tightened and controlled.

The Shainin System is built around a set of tools that are plainly understood and easily applied, hereby refraining from more advanced techniques. The theory is clarified using a clear vocabulary (featuring concepts as 'Red X' and 'Homing in Strategy').

Taguchi's methods

In the Eighties interest in quality improvement among quality engineers and statisticians in the West grew substantially. Most emblematic among the originators of this interest is the Japanese engineer Genichi Taguchi. Taguchi discerns between off-line and on-line quality control. Off-line quality control concerns the design (or re-design) of products and processes, and includes the stages system design, parameter design and tolerance design (Taguchi, 1986, pp. 75–79; see also Kackar, 1985). On-line quality control (during production) has three forms: process diagnosis and adjustment (=*process control*), prediction and correction (=*feedforward* and *feedback control*) and measurement and action (=*inspection*) (Taguchi, 1986, pp. 83–92). Restricted to operational production processes Taguchi's off-line quality control conforms to my definition of quality improvement.

Taguchi invented and promoted various new concepts, such as his quadratic loss function (Taguchi, 1986, p. 15). These concepts represent a different view on quality, in which variation plays a dominant role. This view on quality is generally accepted (Nair, 1992). Furthermore, Taguchi introduced an alternative experimentation methodology (using orthogonal arrays; see Ross, 1988). The adequacy of this methodology has been the subject of much debate among statisticians (Nair, 1992), though it is popular in business practice. As an operationalisation of Taguchi's methodologies and concepts I consider a stepwise strategy described by Ross (1988). This approach is built around Taguchi's quantitative experimentation methodology.

Taguchi emphasises the importance of variation reduction in quality improvement. This can be accomplished in two ways. Based on the results of an experiment settings for the control variables are chosen such that the process is made robust against variation in the nuisance variables. If this action is not sufficient, tolerance design is exploited to accomplish a further reduction in variation.

Taguchi, an engineer himself, uses a vocabulary that is typical for engineers and which differs to some extent from the statistical vocabulary that is used in traditional quality improvement (including terms like 'parameter design' and 'signal-to-noise ratio'). Having a certain degree of refinement without being too mathematical, the methodology should be readily understandable to engineers.

The Six Sigma Programme

Six Sigma is a philosophy for company wide quality improvement. It was developed by Motorola and popularised by General Electric. Several variants are current (compare, for example, the approaches described in Harry, 1997, Breyfogle, 1999, and Pyzdek, 2001). For the strategical and methodological aspects I discuss the variant as presented by Harry (1997), which was introduced at General Electric. For a description of the tools and techniques I consulted Breyfogle (1999).

The programme is characterised by its customer driven approach, by its emphasis on decision making based on quantitative data and by its priority on saving money. The selection of projects is based on these three aspects. Part of the Six Sigma programme is a twelve step 'Breakthrough Cookbook' (Inner MAIC-loop), a problem solving method "specifically designed to lead a Six Sigma Black Belt to significant improvement within a defined process" (Harry, 1997, pp. 21.18–19). It tackles problems in four phases: Measure, Analyze, Improve and Control. The Breakthrough Cookbook is part of an embracing strategy — the Outer MAIC-loop — which comprises the strategical co-

ordination of improvement projects (ibid., pp. 21.21–22). The twelve-step Inner MAIC-loop is studied here as statistical improvement strategy.
The Six Sigma programme is a complete programme for company wide quality improvement, encompassing methods for analysing the customer's demands and for selecting the problems having the highest priority. It features virtually all relevant tools and techniques that have been developed in industrial statistics, from control charting to design of experiments, and from robust design to tolerance design.
The programme is set-up in a way that it can be applied to a range of areas, from manufacturing to services. The implementation and application in the organisation are co-ordinated by 'Champions' and 'Master Black Belts'. Projects are conducted by 'Black Belts' and 'Green Belts', who are selected from middle management.

3.1.2 The collation

I enumerate the elements of the proposed framework and I discuss the corresponding elements in the selected strategies.

Framework	Shainin	Taguchi	Six Sigma
External CTQ	Green Y		Critical To Satisfaction (CTS)
Internal CTQ		(Loss, Signal-to-noise ratio)	Critical To Quality, Delivery or Cost (CTQ, CTD, CTC)
Influence factor	Cause, X	Factor	Cause, X, leverage variable, source of variation
Control variable		Control factor	Controllable factor
Nuisance variable		Noise factor	Uncontrollable factor
Disturbance			

Table 3.1 Comparison of strategies on **C1** and **C3**.

C1: The concepts CTQ and influence factor, and
C3: The concepts control variable, nuisance variable and disturbance
Table 3.1 gives an overview of the equivalent concepts in the discussed strategies. In the Shainin System, a specific quality characteristic that is important to the customer is called Green Y. For its influence factors — the 'causes' or Xs — Shainin stresses the Pareto principle, which sifts the vital few causes (the Red X, Pink X, Pale Pink X, et cetera) from the trivial many (R. Shainin, 1993). Shainin presents no distinction of causes which is comparable to **C3**.
Projects following Taguchi's approach focus on the loss of poor quality, rather than on a quality characteristic. As a consequence, Taguchi's methodology does not offer an equivalent for the concept CTQ. Experiments and analyses focus on the loss function $L(y)$, which represents the monetary loss that an arbitrary customer is likely to suffer

as a function of a quality characteristic y. Taguchi (1986, p. 15) motivates that $L(y)$ can be approximated by a quadratic function having a minimum in t, the target value of y. The loss function is estimated by a series of performance metrics called 'signal-to-noise ratios' (see Ross, 1988, ch. 8-4; or León, Shoemaker and Kacker, 1987). These metrics are taken as the responses in experimentation.
The distinction between control variables ('control factors' in Taguchi's terminology) and nuisance variables ('noise factors') plays an important role in Taguchi's parameter design. Taguchi lists three categories of noise factors (Taguchi, 1986, p. 73):

1. External noise: variables in the environment or conditions of use that disturb the functions of a product.
2. Deterioration (internal) noise: changes that occur when a product deteriorates during storage or wears out during use.
3. Variational (unit-to-unit) noise: differences between individual products that are manufactured to the same specifications.

In the Six Sigma programme, the needs of the customer are translated into critical-to-satisfaction (CTS) characteristics — our external CTQs. These are related to characteristics which are critical to quality, delivery or cost (CTQ, CTD, CTC) (Harry, 1997, p. 12.20). Influence factors are referred to under a variety of names, such as causes, Xs, leverage variables and sources of variation. Breyfogle (1999) introduces the terms Key Process Output Variable (KPOV) and Key Process Input Variable (KPIV) for CTQ and influence factor. Copying Taguchi's approach to parameter design, Breyfogle (1999, ch. 32) introduces the distinction between control and nuisance variables ('controllable' and 'uncontrollable factors' in his terminology). Disturbances are not explicitly distinguished in either Taguchi's methodology or the Six Sigma programme.

C2:. The concepts Operationalisation, Exploration, Elaboration, Confirmation and Conclusion, and

H2: The heuristic that the phases in this order are a promising approach

Table 3.2 compares the proposed phases and their order. The steps listed under the Shainin System were extracted from R. Shainin (1993, figure 2). Taking the sixteen "steps in experimentation" listed in Ross (1988, pp. 203–205) to represent Taguchi's step plan, I notice a strong emphasis on the experimentation phases Elaboration and Confirmation.
The Six Sigma programme groups its twelve steps in four phases (Harry, 1997, pp. 21.18–19):

- *Measurement*: a product related critical-to-quality (CTQ) characteristic is targeted and its performance on the 'sigma scale' of quality defined.
- *Analysis*: the principal sources of variation in the CTQ are identified.
- *Improvement*: the 'vital few' variables which govern the CTQ's performance are surfaced and with this knowledge operating limits for the leverage variables can be established.
- *Control*: a control scheme is identified and deployed for the vital few variables.

These descriptions match to a large extent the functions of the phases Operationalisation, Exploration, Confirmation and Conclusion respectively. However, the division

Framework	Shainin	Taguchi	Six Sigma
Operationalisation	1. Define the project. 2. Establish effective measuring system.	1. State the problem to be solved. 2. Determine the objective of the experiment: • identify the performance metric. • identify the level of performance when the experiment is complete. 3. Determine the measurement method(s).	1. Select the CTQ characteristic. 2. Define performance standards. 3. Validate measurement system. 4. Establish product capability. 5. Define performance objectives.
Exploration	3. Generate clues.	4. Identify factors which are believed to influence the performance characteristic(s)	6. Identify variation sources.
Elaboration	4. List suspect variables.	5. Separate factors into control and noise factors. 6. Determine the number of levels and values for all factors. 7. Identify control factors that may interact.	7. Screen potential causes. 8. Discover variable relationships.
Confirmation	5. Statistically designed experiment.	8. Draw the required linear graph. 9. Select orthogonal arrays. 10. Assign factors and interactions to columns. 11. Conduct the experiment. 12. Analyze the data. 13. Interpret the results.	
Conclusion	6. Return to 3. if Red X not found. 7. Optimize interaction. 8. Realistic tolerances. 9. Irreversible corrective action. 10. Statistical process control. 11. Monitor results.	14. Select optimum levels of most influential control factors and predict expected results. 15. Run a confirmation experiment. 16. Return to step 4. if objective is not met.	9. Establish operating tolerance. 10. Validate measurement system (Xs). 11. Determine process capability. 12. Implement process controls.

Table 3.2 Comparison of strategies on **C2** and **H2**.

of the twelve steps of the Breakthrough Cookbook over Harry's four phases deviates from the grouping in table 3.2, in that steps 4. and 5. are grouped under Analysis, and step 9. under Improvement. This seems dictated more by the desire to have three steps in each phase than by methodological arguments.
The steps of the discussed strategies fit — in their original order — well in the proposed phases. The Elaboration phase cannot be clearly distinguished in the discussed strategies. In Taguchi's methodology no clear delimitation between the Elaboration and the Confirmation phase could be found. The Six Sigma programme does not list actions such as defined in section 2.2.5 explicitly.

H1: The heuristic that improvement actions are derived from found relations in the process

This basic approach is followed in all selected strategies. In the Shainin System the dominant influence factors are identified, which enables the inquirer to explain the behaviour of Y, predict the behaviour of Y, and therefore compute adequate tolerances for the Xs. Therefore, the basic approach follows explain → predict → control.
Also in Taguchi's methodology improvement actions are based on modelled relations

between factors in the process and performance characteristics. However, the focus is on improvement rather than on explanation and understanding, which is reflected in that expected quadratic loss is modelled instead of a CTQ itself. In Nair (1992) Shin Taguchi proclaims:

> Notice that the objective of parameter design is very different from a pure scientific study. (..) Pure science strives to discover the causal relationships and to understand the mechanics of how things happen. Engineering, however, strives to achieve the result needed to satisfy the customer.

In the same paper, Box, among others, declares his profound disagreement with this claim.

The search for causal relationships is omnipresent in Six Sigma: "Supporting the approach is the central belief that the product is a function of the design and the manufacturing process which must produce it" (Harry, 1997, p. 21.2). This is symbolised as $Y = f(X)$, where Y is characterised as dependent, output, effect, symptom, and its role as 'to be monitored'. The X is described as independent, input, cause, problem, and its role as 'to be controlled' (ibid., p. 3.9). The view is that the emphasis should shift from monitoring Y to controlling the relevant Xs (ibid., p. 12.24).

H3: The heuristics provided for the discovery of possible influence factors

Several heuristics were provided in section 2.2.4 to make the discovery of influence factors more efficient. The Shainin System advocates the zooming-in strategy explicitly and provides two techniques which make use of it: the multi-vari chart and paired comparisons (see Bhote, 1991, chs. 6 and 8). The multi-vari chart, which is not only inspected to zoom-in on the relevant class of causes, is also inspected for patterns that could reveal a possible influence factor, and thus provides an example of the application of exploratory data analysis techniques in the Shainin System.

Shainin rejects qualitative investigation for the work in the Exploration phase, to the favour of quantitative inductive techniques: "There is no place for subjective methods such as brainstorming or fish bone diagrams in serious technical problem solving" (P. Shainin, 1993). Shainin's emphasis that hypothesised influence factors should be derived inductively from quantitative observations should not be seen in the light of justification, but in the light of discovery. The claim is that this approach is more effective, although it is not clear whether it is claimed to be effective in comparison with qualitative methods only, or in comparison with the combination of qualitative and quantitative methods (as is advocated in the proposed framework). This empirical claim should be substantiated with empirical evidence, but the only argumentation that is provided is that the strategy eliminates large groups of causes at once (following the zooming-in strategy as described in section 2.2.4). This argument holds only for the approach used in the Exploration and Elaboration phases (in the Confirmation phase possible influence factors are tested one-by-one until the Red X is found — no elimination of groups of Xs here).

The claimed effectiveness could imply two statements, both of which I discuss.

1. *Inductive tools succeed more often in including the important influence factors among the possible influence factors than qualitative tools* (or. than inductive tools combined with qualitative tools). This claim should be substantiated with empirical evidence and seems highly situation dependent. Moreover, many influence factors

do not show in data, such as control variables which are never changed. These cannot but be identified from knowledge and experience.

2. *Quantitative evidence is more effective in prioritising possible influence factors for testing than knowledge and experience* (or: than knowledge and experience combined with quantitative evidence). This claim seems in contradiction with the fact that the effectiveness of an important testing procedure in the Shainin System — variables search — depends completely on the ability of the inquirer to assess correctly the importance of each of the possible Xs (see Ledolter and Swersey, 1997).

I conclude that Shainin deserves credits for emphasising the importance of effective quantitative procedures for discovery, but takes a position which is too rigid to hold in general.

In Taguchi's methods the identification of possible influence factors is limited to the basic tools of brainstorming, flowcharting and fishbone charting (Ross, 1988, section 3-4-1). In the Six Sigma programme a vast collection of tools and techniques is suggested (flowcharting, brainstorming, cause and effect diagrams, run charts, control charts, multi-vari charts, et cetera. See, e.g., Breyfogle, 1999, chs. 4, 5 and 15). However, these techniques are not placed in a strategy, heuristic, or other method.

More research is needed after approaches for discovery and exploration, not focussing on new tools, but studying the underlying heuristics.

H4: The heuristic that learning by scientific method is intrinsically iterative

Taguchi has been criticised for not recognising the sequential and iterative nature of learning. In Nair (1992), for instance, Box criticises Taguchi for being "(..) intended only to pick the 'optimum' factor combination from a *one-shot experiment*" (emphasis is mine). In the same article, Myers and Vining express a similar criticism.

Neither do the Shainin System and the Six Sigma programme emphasise the iterative nature of learning. The notions of learning from error and that hypotheses and even the problem definition can be modified when insight advances are completely absent. I consider this an important failure in the strategies that are discussed.

H5: The heuristics provided by the proposed improvement patterns

The Shainin system focuses on problem solving and tolerance design. Robust design (in combination with adjustment of the mean) was introduced by Taguchi under the name 'parameter design'. First, dispersion — as measured by a signal-to-noise ratio — should be minimised, and thereupon the process mean should be brought on target. Only if robust design is not adequate should tolerance design be applied.

Six Sigma centers around experimentation. Among the suggested improvement actions are adjustment of the mean, robust design on a nuisance variable (Breyfogle, 1999, ch. 32), feedback control (briefly discussed in ibid., ch. 36), mistake proofing (ibid., ch. 38), and tolerance design (step 9. in the Breakthrough Cookbook).

Feedforward and feedback control are underemphasised in the discussed strategies.

R1: The methodological rules that all conjectures are tested to empirical data, and
R4: that specifies when additional testing is required (the severity criterion)

All three strategies are testing approaches in which conjectures are subjected to empirical tests before a conclusion is reached. However, none of the three strategies realises that — according to the severity criterion — additional testing is not always required.

R2: The methodological rule that CTQs and influence factors should be defined operationally, and

R3: the rules that specify when a problem is adequately defined

Six Sigma pays adequate attention to the operational definition of CTQs, influence factors and the problem under study. CTQs are made operational in the Measurement phase. An opportunity for nonconformance requires (Harry, 1997, pp. 12.9–10):

- A characteristic: the attribute, trait, property or quality to be measured.
- A scale: the relative basis for measuring a characteristic.
- A standard: the criterion state or condition specifying nonconformance.
- A density: the empirical distribution of the observations made on this characteristic.

The objective of the project — in terms of the chosen metric — is stated in step 5. of the Breakthrough Cookbook. These demands conform closely to the requirements that were stated in section 2.2.3.

In Taguchi's methodology the focus is on the selection of the relevant signal-to-noise ratio, not so much on the precise definition of the problem in the form of a measurable characteristic. The current performance of the process is not assessed, and as a consequence, there is no check that the selected problem and the translation into a performance metric are suitable. The operational definition of influence factors is not dealt with explicitly.

In the literature on the Shainin System I could not find elaborate statements about operational definitions of CTQs, influence factors or the problem. This is a serious shortcoming.

3.1.3 Discussion

In general the selected strategies conform reasonably well with the proposed framework. Where they deviate, I consider the deviations to be shortcomings of the discussed strategies, not of the proposed framework. This assessment is open for criticism.

In the beforementioned paper by De Mast, Schippers, Does and Van den Heuvel (2000) the authors extend the collation to the level of the tools and techniques that are offered by the discussed strategies. Because tools and techniques are not a part of the proposed framework, I leave this material out of the current discussion.

3.2 Two case studies

3.2.1 Control of the moisture content of coffee

In order to make filter grind coffee, coffee beans are roasted, cooled with water ('extinguished') and ground. The ground coffee is then sealed, palletised and shipped. A project was initiated to gain better control of the percentage of moisture in the ground and packed coffee (external CTQ). Below I present the approach that the proposed framework generated for this project. For confidentiality reasons figures are multiplied by a scalar; it should therefore be noted that the actual moisture content of coffee is of a different magnitude than the mentioned figures.

Operationalisation

It was undesirable to measure the moisture percentage of packed coffee. The easiest place in the process to measure the moisture content was after the beans are roasted and before they are ground. Therefore, the internal CTQ was defined as: the percentage of moisture in the coffee beans after the roasting and extinguishing steps. There was a direct link between this moisture percentage and the moisture percentage of the packed coffee, whence control of the internal CTQ would give the desired control of the external CTQ.

The measurement procedure consisted in an operator determining the moisture percentage of a sample of beans. Samples were taken from the first of every 15 batches. The moisture percentage should be below 12.6%, a figure which was based on marketing research and laboratory research after the effect of moisture on the keeping qualities of packed coffee. Given this upper limit, the company's objectives demanded that the average moisture content be as large as possible.

In order to assess the current performance, 1231 measurements were collected (dataset *I*). They can be summarised as follows:

- The mean moisture percentage: 11.0%.
- The short-term standard deviation (which is estimated by our statistical software package on the basis of the average moving range and ignoring outliers): 0.336.
- The long-term standard deviation (estimated as the sample standard deviation): 0.532.

It was stated that the objective of the project was to bring the long-term standard deviation below 0.280, so that the average moisture percentage could be increased, but yet kept at a safe distance from 12.6%. From a gauge repeatability and reproducibility study it appeared that the measurement procedure which was used to measure moisture percentage was too unreliable. An in-line measurement device was installed, which reduced the short-term standard deviation from 0.336 to 0.123 — a dramatic reduction of the variation in the measured moisture content.

Exploration

Inspection of the 1231 measurements of dataset *I* learned that there were differences among the five roasting machines. Furthermore, additional variation between days was observed and regularly there were outliers. A brainstorming — partly inspired by these observations — resulted in the identification of the following potential influence factors:

- Roasting machines (nuisance variable),
- Weather conditions (nuisance variable),
- Stagnations in the transport system (disturbance),
- The size of a batch of coffee in kilogrammes (nuisance variable),
- The amount of water added after the roasting process (control variable).

Elaboration/Confirmation

The differences among the machines were confirmed and estimated using the analysis of variance technique. Furthermore, a 3^2-experiment was conducted to study the effect of the size of a batch of coffee (SB) and the amount of added water (AW) onto the moisture percentage ($M\%$) (dataset *II*). Analysis of the results confirmed the effects of both factors and gave the following model:

$$M\% = -46.40 + 0.064\,SB + 0.70\,AW - 0.0008\,SB \cdot AW. \tag{3.1}$$

Conclusion

A feedforward system was designed in which the amount of added water was used to compensate for differences among machines and variation in batch sizes. The correction for batch sizes follows the formula

$$AW = \frac{tM\% + 46.40 - 0.064\,SB}{0.70 - 0.0008\,SB},$$

which was derived from formula (3.1) ($tM\%$ denotes the target value for moisture percentage).

Elaboration

The conjectured effect of weather conditions was made more specific. It was hypothesised that the relevant characteristic of the weather is the humidity. A device to measure humidity was borrowed from another factory.

Confirmation

Humidity measurements were correlated with the 1009 associated moisture percentages (dataset *III*). The conjectured influence could not be established, leaving the fluctuations in the process unexplained.
From the same measurements it appeared that the improvement actions so far (a better measurement system, feedforward control for differences among machines and variation in batch sizes) had reduced the variation in the moisture percentage to 0.201 (long-term standard deviation) and 0.066 (short-term standard deviation ignoring outliers). The effectiveness of the feedforward systems was confirmed. Moreover, outliers appeared to coincide with stagnation of the transport system, thus confirming this hypothesised influence factor.
In fact the objective of the project was met. It was decided, however, to use the found influence of the amount of added water AW to build a feedback system to compensate for the still unexplained fluctuations in the process.

Conclusion

From the 1009 measurements of dataset *III* 18 apparent outliers were removed. From the remaining 991 data an IMA(1,1) model could be estimated. The theory that describes this kind of modelling can be found in Box, Jenkins and Reinsel (1994). The estimated model is

$$y_t - y_{t-1} = -0.6186\,a_{t-1} + a_t,$$

where y_t is the moisture percentage measured at time t, and the a_t have identical distributions with mean 0 and estimated standard deviation 0.088. A residual plot shows that their distribution has somewhat heavier tails than the normal distribution. The autocorrelation and partial autocorrelation function show that the estimated model

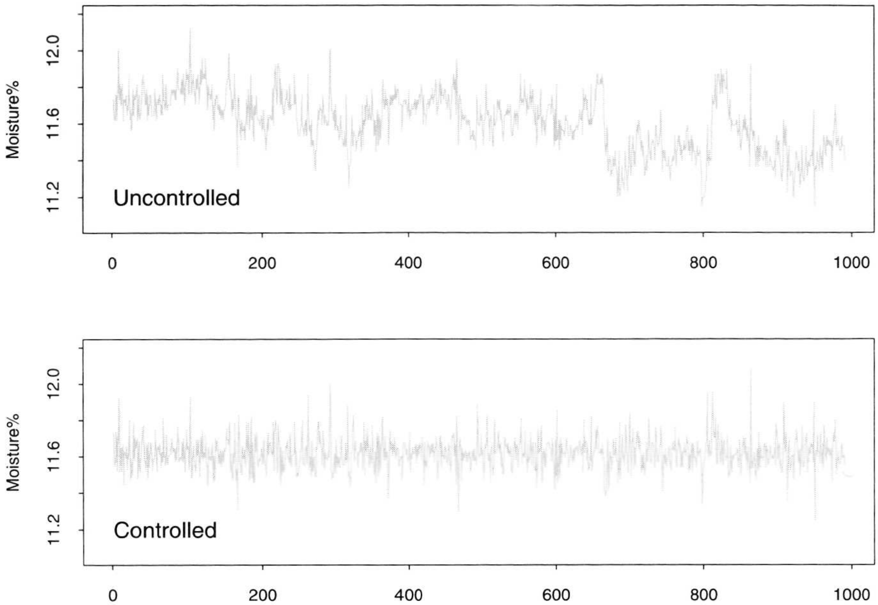

Figure 3.1 Moisture percentages with and without feedback adjustments.

reduces the observations effectively to uncorrelated noise. It can be shown (Box, Jenkins and Reinsel, 1994) that for a process that is well described by the given model the controller that minimises mean square error (the MMSE controller) is given by:

$$AW_t - AW_{t-1} = -Ge_t/g,$$

where at time t: $AW_t - AW_{t-1}$ represents the adjustment of added water, e_t is the observed deviation from target, g is the process gain, which is given by the effect of adding 1 liter of water, and G is the damping factor. For this process, $g = 0.70$ and $G = 1 - 0.6186 = 0.38$. The feedback system reduced the long-term standard deviation to 0.10. The effectiveness of the feedback controller is demonstrated in figure 3.1. The upper graph shows the 991 measurements of dataset *III*. The lower graph shows what these measurements would have been if the feedback controller had adjusted the process.

Conclusion

As a result of our improvement actions management allowed to increase the average moisture percentage to a value safely below 12.6% (adjustment of the mean). An SPC control loop was designed to monitor the behaviour of the feedforward and feedback control systems.

Discussion

Various conditions made this an easy project. Measurements were practically for free and could be obtained in huge numbers. The discovery of influence factors was particularly easy. Consequently, the emphasis in the project was on the Confirmation and

Conclusion phases. If we are prepared to interpret measurement error as a nuisance variable, we see that all improvement actions were derived from found relations in the process. One of the most effective improvement actions — a better measurement system for moisture percentage — was a result not so much of complicated inquiry, but rather of focusing attention of engineers to the issue.
The case illustrates that projects begin following the Operationalisation → Exploration → Elaboration → Confirmation → Conclusion sequence, but somewhere in the project advanced insight makes it necessary to come back to earlier phases.

3.2.2 Reduction of scrap in biscuit production

A certain type of biscuits is produced in three main steps:

1. Preparation of dough;
2. Punching the biscuits out of the dough and baking;
3. Frosting of the baked biscuits.

A large percentage of the baked biscuits was not suitable for processing on the frosting machines and was therefore wasted. An improvement project was initiated to bring down this percentage of scrap. For reasons of confidentiality percentages of scrap or other performance metrics are not reported.

Operationalisation
The external CTQ being suitability of biscuits for processing on the frosting machines, several internal CTQs were considered: the weight, length, width, thickness and hardness of the baked biscuits. For each CTQ the specifications were considered, as well as the measurement procedure. Measurements were collected on all of these CTQs. The performance on hardness was by far the worst (see the results of the capability analysis as presented by Minitab in figure 3.2). Given the lower specification limit of 4.7 Bar, it appeared that many biscuits did not have the required hardness. Length and thickness were as well problematic CTQs.
It is important to realise that the poor performance of the process on the selected CTQs has no consequences for buying customers, but leads to problems on the frosting machines. Although the amounts of scrap that are accepted in this type of processes is larger than in other types, gaining better control over these CTQs was expected to be profitable. Hardness was selected as the main internal CTQ, taking length and thickness along as secondary CTQs.
From a measurement reliability study it appeared that the expensive laboratory measurements on hardness were not more reliable than the measurements performed by the operators themselves. This result made it easier to use hardness measurements in the project and in production.
The current procedures required that each pallet of biscuits would be inspected on length, width and thickness. The procedure required the operator to measure a sample of 14 biscuits (collected from two trays). Based on the results pallets were accepted or (completely) rejected. A variance decomposition of 88 biscuits from various pallets showed that the dominating variance component was between trays, not between pallets. This suggested that it was not wise to reject or accept complete pallets. A more sophisticated inspection scheme was designed. Furthermore, a registration system for

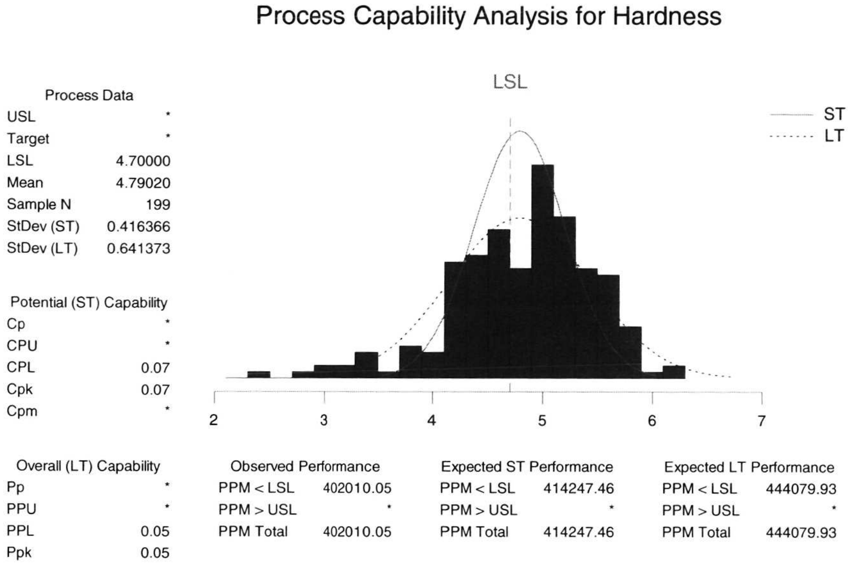

Figure 3.2 Capability analysis for hardness of biscuits.

the results of the inspections was set up so that the information could be used for further improvement projects.

Exploration

The data that were collected in the Operationalisation phase were analysed using control charts and multi-vari charts. The analyses did not result in specific clues. Brainstorming sessions were more successful. These sessions were structured following the steps of the production process and using the standard classes Man, Machine, Method, Material, Measurement and Mother Nature. Besides, a failure mode and effect analysis was performed with the operators, in which recurring disturbances were identified.

Elaboration

The Exploration phase resulted in a process matrix featuring 59 potential influence factors. A part of it — the influence factors related to the preparation of dough — is listed in table 3.3. The presumed effect of many of these influence factors was rationally considered, which resulted in some factors being marked *not promising*. The experiments and other investigations were prioritised and organised. In order to measure some of the influence factors related to dough (such as temperature) it was necessary to perform a homogeneity study. Based on this study it was possible to give an adequate operational definition of factors such as dough temperature.

Conclusion

Many of the identified disturbances could be tackled by composing check lists for maintenance work and for start-ups. Furthermore, on-line inspections of the running process were better organised so that disturbances would be reacted to more alertly (*mistake proofing*).

Influence factor	Type	L	T	H
Dosage of meal	Control variable	×	×	×
Quality of the meal	Nuisance variable	×	×	×
Dosage of Y-fat	Control variable	×	×	×
Quality of Y-fat	Nuisance variable	×	×	
Dosage of sugar	Control variable	×	×	×
Dosage of syrup	Control variable	×	×	×
Dosage of cold water	Control variable	×	×	×
Dosage of hot water	Control variable	×	×	×
Dosage of pyrosulphate	Control variable	×	×	×
Dosage of soda	Control variable	×	×	×
Dosage of crumb	Control variable	×	×	×
Order of the additions	Control variable	×	×	×
Blending degree of pyrosulphate	Nuisance variable	×	×	×
Duration of blending of flour with water	Control variable	×	×	×
Contaminated level detector	Disturbance	×	×	×
Degree of blending of dough	Nuisance variable	×	×	×
Dough temperature	Control variable	×	×	×
Bad raw materials	Disturbance	×	×	×

Table 3.3 Part of the process matrix, showing influence factors for L(ength), T(hickness) and H(ardness).

Confirmation

Before the biscuits are processed on the frosting machine they are set to rest during a number of days. The duration of this period was identified as a potential influence factor (*control variable*) and an investigation confirmed the influence of this duration onto the hardness (0.10 Bar per day). At this point, this effect could, however, not be explained.

The effects of several other possible influence factors were *not* confirmed, which put an end to much debate in the factory. In a 2^3 experiment with 4 runs in the centre point the effects of three parameters of the oven were tested. None of these factors appeared to have any effect on the hardness of the biscuits, much to the surprise of the engineers, and although the settings were varied quite extremely. The factor *ventilator speed*, however, had a small dispersion effect for length and thickness.

Conclusion

The ventilator speed was adjusted slightly to reduce the dispersion of length and thickness of biscuits (*robust design on dispersion effect*).

Confirmation

In order to study the potential influence factors related to dough preparation, we conducted a 2^{5-1} experiment with 4 runs in the centre point and in two blocks. This design has resolution V, but the block effect is confounded with a two factor interaction. It was particularly difficult to control all conditions, such as the time between dough preparation and dough processing, or the sticking of the biscuits to the oven

belt. Many operators and engineers were called in to help keeping conditions as constant as possible. A second complication was that a change in some of the factors had no immediate effect, but required some 40 minutes for stabilisation.
Analysis learned that one of the factors had no influence. Some of the others had quadratic effects, whence additional measurements had to be collected following the runs of a central composite design. From the analysis optimal settings could be found for the factors, but the gain in hardness was not gigantic: 0.5 Bar. Moreover, the required change in recipe came across resistance from the marketing department.
The way the biscuits are stacked after they are baked (wide apart or tightly) appeared to have an important influence on hardness. Moreover, the biscuits produced during the beforementioned experiments were remeasured after several days. In this period their hardness had increased by around 0.1 Bar per day.

Elaboration
At this moment the situation was that those factors that had been expected to have important influence onto the hardness appeared not or hardly relevant. On the other hand: when the biscuits were left to rest, their hardness increased (especially when they were kept wide apart), and this effect was larger than the effects of all other factors studied so far. Reasoning resulted in the conjecture that the observed effect could be due to humidity.

Confirmation
We followed some trays of biscuits during a period of 13 days. Some of these trays were left open, others were closed, whereas the biscuits in two trays were sealed in small vacuum bags. The found differences were large: after just 4 days the biscuits in the open trays were 1.5 Bar harder than the biscuits in the closed trays and the sealed biscuits. The moisture content of the biscuits was measured and correlated with their hardness. This variable appeared to be the explanatory variable that was searched for. The estimated relation was

$$Hardness = 4.65 + 0.38\,Moisture\%.$$

Highly unexpected as this result was, it was unclear how the discovered effect could be put to use.
The effect of the biscuits' sticking to the oven belt (*nuisance variable*) was studied, and the influence appeared huge. Even when the biscuits stuck slightly to the belt, the hardness was affected, but the effect could increase up to 1.5 Bar.

Conclusion
Several countermeasures against sticking were tried (scattering of flour underneath the biscuits and the cleaning of the oven belt during production) before a successful countermeasure was found: a certain type of grease on the oven belt resolved the problem. The impact of this improvement was huge.
The quality control system was reconsidered in order to incorporate the changes in the process. Moreover, the control system could be made more effective now that new knowledge about the process was gathered.

Discussion
The contrast between this project and the one that was the subject of the first case study

is large. For several reasons this was a difficult project:

- The process was very complex, and many influence factors were important.
- Many influence factors were hard to control. Especially properties of dough, which is a highly unstable material, appeared to live their own life.
- Measurements had to be done by hand and were labour-intensive.
- Much about the process was unknown (for example: realistic requirements on such CTQs as hardness, length and thickness), but figuring everything out was simply too much work.
- Operators working in the baking and the frosting processes sometimes had opposite interests.

The benefits of the project were more than the prevented disturbances, the solved problem of the biscuits sticking to the oven belt, the optimised settings for the oven and dough preparation, and the improved inspection system. The effect of moisture, though not directly applicable, could explain many strange observations. Moreover, many myths were invalidated, thus focusing future inquiries.

The process was far from the state of statistical control and Deming would probably have insisted that it should be brought in that state before improvement could be strived after (see the appendix to chapter 1 of this thesis). It is, however, unrealistic to suppose that this process will ever approach the state of statistical control, because:

- Dough is a highly unstable substance, whose structure changes from minute to minute and is highly dependent on temperature and other influences.
- The process works with natural materials such as flour, the variability of which is unavoidable.
- The process is so complex and so many things can go wrong, that inevitably some will go wrong each day.

This situation made experimentation certainly more complicated, but not at all impossible. My assessment is that Shewhart's (1939, p. 25) procedure alone — continuously reacting to signals of assignable causes, and then discovering their nature and removing them — is not effective for improving this process. There are so many signals of assignable causes that a search should not be instigated reactively, but in a structured and organised manner. Moreover, the discovery and removal of assignable causes is in many cases complicated. Taken together, the situation required the organised, systematic and off-line type of inquiry that I call *improvement project*, instead of Shewhart's reactive quality control procedure.

A very important point is, though, that an effective use of the results of a quality improvement project requires that full use is made of the found knowledge to improve the quality control system. The case study illustrates this point: due to practical circumstances the quality control system was at first not adequately improved, leading to a month of bad results. Thereupon, the control system was upgraded in order to take full advantage of the results of the project. The percentage of scrap was subsequently reduced by a considerable amount.

3.3 Further experiences in consultancy work

The formulation of the proposed framework was — of course — an iterative procedure in which proposed elements were critically confronted over and over again with other material. An important role was played in this context by the critical confrontation of proposed theory with my consultancy work in quality improvement.
The results of the confrontation of earlier versions of the proposed framework with my consultancy experience were incorporated in the final version as described in chapter 2. Some observations, however, were quite illuminating and these are discussed in this section.

3.3.1 Canonical structures of the explanatory network

In some projects the explanatory network has a typical form. The identification of these typical forms enables a categorisation of improvement projects. An interesting example is provided by a project in a caffeine extraction process. The objective of the project was stated to be the increase of the amount of decaffeinated coffee per week (measured in tonnes). This external CTQ was associated with the internal CTQ *cycle time*. The desired decrease in cycle time could not simply be brought about by the reduction of the total duration of the extractions, because an important restriction was that the caffeine percentage of the decaffeinated coffee should be below 0.1%.
I have encountered similar situations, in which the objective is to increase capacity or reduce cycle time, but where an important restrictive variable is in play. Typically in this kind of projects, the influence factors for the capacity and their effect are quite easily found. Rather, the project centers around the identification of the influence factors for the restrictive variable (or even the restrictive variables themselves). The idea is that an increase in capacity can be arrived at by a better control of the restrictive variable.

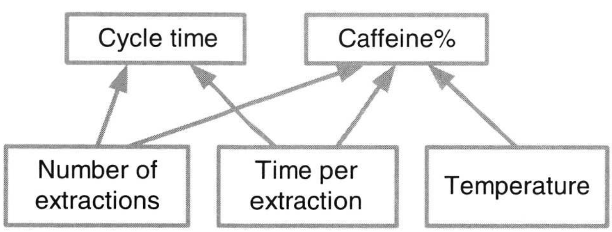

Figure 3.3 CTQs and influence factors for the caffeine extraction process.

The explanatory network for the project in the caffeine extraction process contains two CTQs and three influence factors (figure 3.3). *Number of extractions* (*NE*) and *time per extraction* (*TE*) cannot straightforwardly be reduced, since this would result in too large a caffeine percentage. However, optimising the setting of *temperature* (*T*) we reduce the caffeine percentage, allowing a reduction in *NE* and *TE*, thus reducing cycle time. Moreover, the effect of *NE* and *TE* onto cycle time is not a multiple of their effect onto caffeine%. This gives an opportunity to find optimal values for *NE* and *TE* under the restriction of keeping caffeine% constant.

The estimated models are:

$$Cycle\ time = TE \cdot NE,$$

$$Caffeine\% = 3.2916 - 0.0638\,TE - 0.0846\,NE - 0.0272\,T + 0.00144\,TE \cdot NE + 0.000566\,TE \cdot T.$$

In the relevant range, the optimal setting for T is 75°C. The overlaid contour plots of predicted cycle time and caffeine% as functions of *NE* and *TE* (T set at 75°C) are given in figure 3.4.

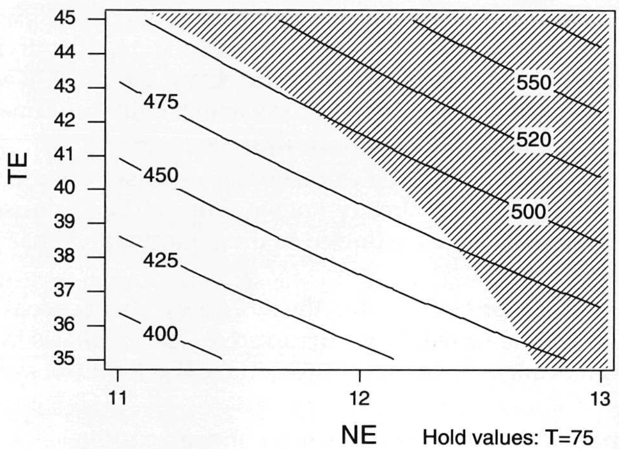

Figure 3.4 Overlaid contour plots for cycle time and caffeine%.

The contour lines represent predicted values of cycle time. The shaded area represents the process settings for which the caffeine percentage is below 0.07%, which is at a safe distance from the specification 0.1%. From the contour plot we choose the settings *NE* = 13, *TE* = 35 minutes, which gives a cycle time of 455 minutes. Compared with the original settings *NE* = 12, *TE* = 45, this gives a reduction of cycle time of $540 - 455 = 85$ minutes.

Interesting research could be done to identify more standard patterns in explanatory networks and related approaches, thus arriving at a classification of improvement projects.

3.3.2 Emphasis in projects

There is hardly an improvement project in which all phases — Operationalisation, Exploration, Elaboration, Confirmation and Conclusion — are equally important. In fact, in many projects the emphasis is on a single or two phases. Some examples.

- *Operationalisation phase.* The objective of a certain project was to reduce dead inventory (inventory for which there is no purpose in the company). Having made this objective more specific, and having collected information on the current amount of dead inventory, the root of the problem appeared already at hand: an inadequate order procedure.

In a process making parboiled rice, one CTQ was the capacity of the installation, under the restriction that another CTQ, namely the extent to which the rice was cooked enough, remained acceptable. The development of an adequate measurement procedure to determine whether rice is cooked enough was an important part of the project. Once such a procedure was found — the final procedure involved counting the number of grains of rice not cooked enough — effective experiments could be designed to establish how much the capacity could be increased without compromising the quality of the cooked rice.

- *Exploration phase.* The CTQ in a project concerning a coffee packaging process was the event that a package of coffee is not vacuum. Identifying the cause of this problem took a lot of detective work. Once the cause was tracked down (insufficient capacity of the pump in the vacuum bell), confirmation of its effect and finding countermeasures were easily done.
- *Confirmation phase.* In a caffeine extraction process (see the preceding example) the influence factors were already known, but not their precise effects. Complicated experiments were conducted in the Confirmation phase to model these effects.
- *Conclusion phase.* In order to control the moisture content of roasted coffee beans the influence factors were quickly discovered. The emphasis in the project was on the design of adequate feedforward and feedback control systems.

3.3.3 Importance of the Operationalisation phase

Upon identifying a CTQ and measuring its current performance, an inquirer not seldomly finds out that the selected CTQ fails to represent the problem he is considering. In a biscuit factory, for instance, an improvement project was started that aimed at (among other things) the reduction of the percentage of biscuits that are too long. Having defined a measurement and sampling procedure for this CTQ, the inquirer collected a number of measurements to assess the current extent of the problem. From the analysis the inquirer found out that:

1. There were indeed many biscuits that were too long, but in addition,
2. with the biscuits that were too long a considerable amount of good biscuits was wasted as well, which was due to the inspection procedure that was employed.

Thus, it appeared that the perceived problem consisted in fact of two problems. I have encountered the situation quite frequently, that the *perceived* problem is not (or only part of) the real problem. In fact, the operationalisation of the problem should be considered an iterative procedure. A definition of a CTQ should be considered as a hypothesis, which can be modified or even rejected on the basis of empirical evidence collected in the Operationalisation phase or in later phases. In my experience, measurement procedures, sampling plans, specifications, and even the characteristic that is assumed to represent the problem are frequently reconsidered later in an improvement project.
From this example I would like to conclude that:

1. The importance of the Operationalisation phase is that the problem is made explicit, but that

2. The results should be considered tentative and open for revision.

3.3.4 Tendency of engineers to be too critical in the Exploration phase

Often, I have encountered a considerable amount of scepticism with engineers when they could not explain or understand from their technical knowledge the potential influence factors that were identified during the Exploration phase. If, however, inquirers do not make bold conjectures, but only pose truisms, they do not learn anything (cf. the quote from Popper in section 2.3). The scepticism is not justified: the hypothetico-deductive method has a safety net in that wrong hypotheses are refuted during the Confirmation phase. Engineers should be convinced to feel safe to explore new ideas, also bold ones, because this is the way breakthroughs are arrived at.

3.3.5 Operational definitions in quality improvement

In quality improvement CTQs are usually borrowed from ordinary language. In order to enable empirical testing and prediction it is vital that they are defined operationally. A standard example with which I confront my students is the problem that in a particular year 17% of the trains of the Dutch railway company was delayed (according to figures of the railway company). Before the delay of trains can be the subject of a quality improvement project the event *train is delayed* should be given operational meaning. This would require at least a specification of:

- A margin: after how many minutes is a train considered late? The railway company specifies: after 3 minutes.
- The time that is measured: time of arrival or time of departure? The railway company specifies: time of arrival.
- The unit of measurement: the statement '17% of the trains is delayed' suggests that the unit is a train. It should then be specified which arrival time of a train is considered (the arrival at the terminal station?). In fact, the unit that the railway company defines is not a train, but an arrival of an arbitrary train at an arbitrary station. Thus, a more de facto statement would be: '17% of the arrivals of trains is overdue'.
- The sampling system: are all arrivals in the Netherlands recorded? The railway company specifies: only the arrivals at the main stations (which is a specified selection of stations) are recorded.

The importance of precise specifications is illustrated from the fact that the consumers' association has deviating figures. Surely, this is a result from the fact that their definition deviates on at least one of the four points mentioned above. This does not mean that the definition of one is valid and the other's is not. As long as the definitions correlate with the external CTQ (which is passenger dissatisfaction due to delays) they are both valid.

An example from my own consultancy work is the CTQ *hardness of biscuits*. This concept is borrowed from ordinary language, but upon applying it to biscuits it appears too simple to be of use. In particular, different (non-correlating) operational definitions are possible (related to tensile strength and success in withstanding pressure exerted in various ways). It is important to select carefully a definition that relates to the external CTQ.

In the same project, the temperature of an oven was identified as potential influence factor. In the oven the temperature was measured at various places and it was not clear which of these temperature measurements were appropriate. The problem was solved by doing the analyses on all of these measurements. The conclusions were identical for different measurements, and this fact gave confidence in the results.