Taylor & Francis
Taylor & Francis Group

An Overview of the Shainin System™ for Quality Improvement

Stefan H. Steiner[1],
R. Jock MacKay[1],
John S. Ramberg[2]

[1]Business and Industrial Statistics
Research Group (BISRG),
Department of Statistics and
Actuarial Sciences, University
of Waterloo, Waterloo, Canada
[2]Professor Emeritus, University of
Arizona and Consultant, Pagosa
Springs, CO, USA

ABSTRACT The Shainin System™ (SS) is the name given to a problem solving system, with its associated strategies and tools, developed by Dorian Shainin, and widely used and promoted in the manufacturing sector. Dorian Shainin also called this system Statistical Engineering, reflecting his engineering education and background. The consulting firm, Shainin LLC, offers the system under the trademarked name Red X® Strategy.

The goal of this article is to provide an overview of SS, a critical assessment, and a brief comparison with other industrial problem solving systems. The emphasis is on a discussion of the guiding philosophy and principles. Some specific SS tools are examined and compared with alternative methods. In our assessment, the Shainin System is valuable for many types of problems and many of its elements have been, or should be, incorporated into other process improvement methodologies.

KEYWORDS B vs. C™, Components Search™, group comparison, Isoplot®, multivari chart, precontrol, progressive search, red X® strategy, Shainin, Six Sigma, Variable Search™

INTRODUCTION

The goal of this article is to provide a critical overview of the Shainin System™ (SS) for quality improvement, developed over many years under the leadership of the late Dorian Shainin. SS is also called Statistical Engineering by the consulting firm Shainin LLC that holds the trademark and Red X® strategy in parts of the automotive sector where SS is popular. The overall methodology has not been subject to critical review although some of the components have been discussed extensively. Here we provide such a review and also compare the Shainin System to other process improvement systems including Six Sigma. We also describe a few of the more controversial and widely used SS statistical methods.

The Shainin System was developed for and is best suited to problem solving on operating, medium to high volume processes where data are cheaply available, statistical methods are widely used and intervention into the

Address correspondence to Stefan H.
Steiner, Department of Statistics and
Actuarial Science, University of
Waterloo, Waterloo, ON, Canada N2L
361. E-mail: shsteiner@waterloo.ca

process is difficult. To our knowledge, it has been mostly applied in parts and assembly operations. We agree with reviewers (Nelson, 1991; Moore, 1993; Hockman, 1994; and Ziegel, 2001) that these books make many unsubstantiated, exaggerated claims. What is worse, we believe that these books are a disservice to SS, since the hyperbole hides many of the genuinely useful ideas. A less technical and less controversial reference that includes many case studies is Traver (1995). Overviews of SS have been published in conference proceedings; see Shainin (1992, 1992b, 1993, 1993b, 1995), Shainin and Shainin (1990), and Shainin et al. (1997). Other review articles include Logothetis (1990) and De Mast et al. (2000). Does et al. (1999) cover many of the specific tools associated with the Shainin System but not the overall strategy. Ledolter and Swersey (1997a, 1997b) review two widely heralded SS tools, precontrol and variables search. There may be new developments not yet in the public domain. Steiner and MacKay (2005) present a variation reduction algorithm that builds on what we think are the best elements of SS.

In assessing the Shainin System, it is important to differentiate between the overall approach that we think is strong, and the specific analysis methods some of which are weak. The article is divided into two major parts. First, we discuss the basic principles underlying SS, and the consequences of applying these principles within the Shainin System. It is the use of these principles and the corresponding algorithm in combination that defines and distinguishes the overall strategy of SS from other approaches. Next, we discuss a selection of SS statistical tools used within the algorithm. By "tool", we mean the data collection plan and the subsequent analysis method. We discuss alternatives to the analysis methods where appropriate.

THE GUIDING PRINCIPLES OF THE SHAININ SYSTEM

We consider the underlying principles of SS in two groups. The first group follows from the idea that there   . This idea
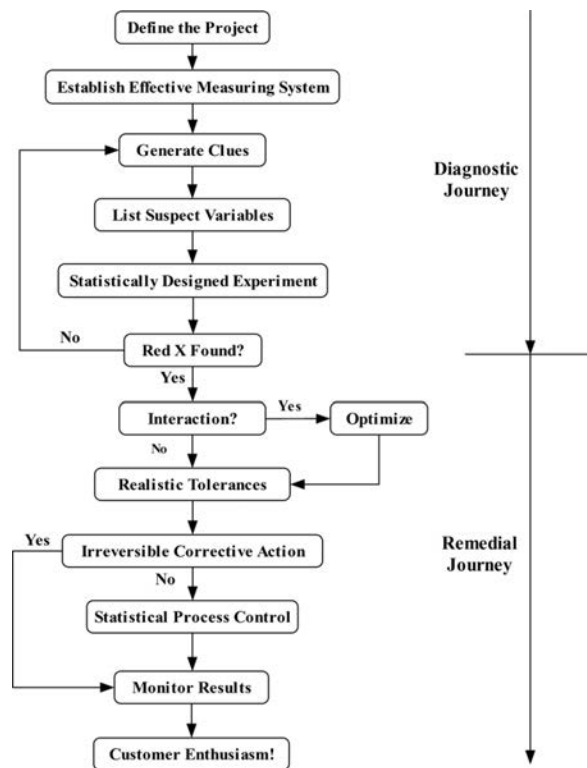
appears in Juran and Gryna (1980), but it is   . The second group of principles is embedded in the algorithm, the Shainin System^{TM}, shown in Figure 1.

DOMINANT CAUSES OF VARIATION AND PROGRESSIVE SEARCH

A fundamental tenet of SS is that, in any problem,   . This presumption is based on an application of the Pareto principle to the causes of variation. Juran and Gryna (1980, p. 105) define a dominant cause as   " In   The emphasis on a dominant cause is justified since "The impact of the Red X is magnified because the combined effect of multiple inputs is calculated as the square root of the sum of squares" (Shainin, 1995). To clarify, if the effects of causes (i.e., process inputs that vary from unit to unit or time to time) are independent and roughly additive, we can

decompose the standard deviation of the output that defines the problem as:

$$\text{stdev(output)} = \sqrt{\begin{array}{l}(\text{stdev due to causes 1})^2 \\ + (\text{stdev due to cause 2})^2 + \ldots\end{array}} \quad (1)$$

A direct consequence of (1) is that we cannot reduce the output standard deviation substantially by identifying and removing or reducing the contribution of a single cause, unless that cause has a large effect. For example, if (stdev due to cause 1) is 30% of the stdev (output), we can reduce the stdev (output) by only about 5% with complete elimination of the contribution of this cause. The assumption that there is a dominant cause (possibly because of an interaction between two or more varying process inputs) is unique to SS and has several consequences in its application.

Within SS, there is recognition that there may be a second or third large cause, called the Pink X^{TM} and Pale Pink X^{TM} respectively (Shainin, 1993b), that make a substantial contribution to the overall variation and must be dealt with in order to solve the problem. Note that if there is not a single dominant cause, reducing variation is much more difficult, since, in light of (1), several large causes would have to be addressed to substantially reduce the overall output variation. To simplify the language, we refer to a dominant cause of the problem, recognizing that there may be more than one important cause.

There is a risk that multiple failure modes contribute to a problem, and hence result in different dominant causes for each mode. In one application, a team used SS to reduce the frequency of leaks in cast iron engine blocks. They made little progress until they realized that there were three categories of leaks, defined by location within the block. When they considered leaks at each location as separate problems, they rapidly determined a dominant cause and a remedy for each problem.

SS uses a process of elimination (Shainin, 1993b), called progressive search, to identify the dominant causes. Progressive search works much like a successful strategy in the game "20 questions," where we attempt to find the correct answer using a series of (yes/no) questions that divide the search space

into smaller and smaller regions. To implement the process of elimination, SS uses families of causes of variation. A family of variation is a group of varying process inputs that act at the same location or in the same time span. Common families include within-part, part-to-part (consecutive), hour-to-hour, day-to-day, cavity-to-cavity and machine-to-machine. At any point in the search, the idea is to divide the inputs remaining as possible dominant causes into mutually exclusive families, and then to carry out an investigation that will eliminate all but one family as the home of the dominant cause.

For example, in a multivari study (see the next section), suppose we find that variation part-to-part is much larger than variation time-to-time. Then, all varying inputs that change over the longer time frame, such as properties of batches of raw material, can be eliminated as dominant causes.

Another consequence of the assumption of a dominant cause is that we can gain a lot of information about this cause by comparing units with extreme values of the output. To our knowledge, this explicit use of "leveraging" is unique to SS. Shainin et al. (1997) refer to comparing the "best of the best" (BOB) and "worst of the worst" (WOW) units. The values of the dominant cause must be substantially different on these two groups of units and hence it will be identifiable. One advantage of leveraging is that we can eliminate families of causes using investigations with small samples of extreme units. The idea of leveraging is specifically employed in many SS tools, including Component Search^{TM}, Variable Search^{TM} and group comparison, discussed later in this article. Note, however, to find a small number of extreme units, we may need to measure the output on a large number of units. Also, the terminology can cause confusion. For outputs with two sided specifications, none of the extreme units is best of the best.

SS shuns brainstorming and cause-and-effect diagrams when screening possible causes. Using cause-and-effect analysis, once all possibilities are identified, we are forced to look at a large number

of potential dominant causes one-at-a-time or in some combination. Using progressive search and carefully designed observational investigations, we can rule out large families without ever identifying the individual varying inputs that make up the family. In our experience, progressive search is much more efficient than brainstorming. Shainin (1993) states, ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮ We agree with this statement when the goal is to find a dominant cause; however, we disagree when we are looking for a solution, having identified a dominant cause.

The success of progressive search depends on our ability to combine empirical knowledge provided by process investigations and engineering/process knowledge. The need for data from the process is emphasized throughout the SS methodology. This emphasis has perhaps led to the misunderstanding (De Mast et al., 2000) that qualitative process knowledge is not required in SS. A team must have deep understanding of the process to construct appropriate families, plan investigations and identify a particular family as the home of a dominant cause. Process knowledge is also essential when determining an appropriate change to the product, process, or control plan that will reduce or eliminate the effect of an identified dominant cause. The necessity for process knowledge is acknowledged in all process improvement systems. However, in SS, there is an increased awareness that in order to progress, engineering process knowledge must be combined with empirical knowledge gained by studying the process.

Within SS, there is no explicit consideration of whether the dominant or any other causes are common or special. The search strategy is designed to look for one or two causes with large effects. For variation reduction problems, using families of variation and the method of elimination is a more effective way to partition the causes than is the classical Statistical Process Control (SPC) division into common and special causes. To substantially reduce the process variation in many cases, we need to address common cause variation and the control chart will be no help in identifying such causes. For example, if the dominant cause acts part to part, then the common cause variation within any time based subgroup will be large and the control chart

will not signal the action of the dominant cause. With families defined specifically based on existing process knowledge, there is a broad array of sampling plans and analysis methods, other than control charts, that can be used to eliminate families.

The focus on finding and eliminating the effects of a dominant cause is appropriate in many problems, but can be restrictive. ▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ (Steiner and MacKay, 1997–1998, 2005).

The use of progressive search is not without difficulties. It can be hard to identify a dominant cause that is an interaction between varying inputs in different families. Progressive search requires patience since multiple investigations are usually required to isolate dominant causes, and it is innately sequential which can be a hard sell in today's fast paced industrial environment. With small sample sizes and the emphasis on extremes, there is a risk of focusing on outliers that are not due to the dominant cause driving the overall variation. The definition of "extreme" requires care, especially in the case of two-sided specifications. It is possible that different dominant causes are responsible for units with output values on opposite sides of the target. ▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ For an artificial example, suppose that there are 20 causes, all of which contribute independently and additively to the variation so that (1) applies. Then, if all causes are equally important, completely removing the effects of half the causes only reduces the output standard deviation by about 30%. In these instances, all problem-solving systems based on Juran's diagnostic and remedial journey will have difficulty because the effect of any one cause is masked by the variation due to all others.

THE PROBLEM SOLVING ALGORITHM

The SS steps for problem solving are given in Figure 1. Note that the algorithm is defined for a single project, and is designed to fit into a larger project selection and management process, not discussed here. See Shainin et al. (1997). ▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

terminology from Juran and Gryna (1980, p. 104). In the diagnostic journey, the problem is defined, the measurement system is assessed, and the dominant cause of variation is identified and verified. In the remedial journey, the effect of the dominant cause is eliminated or reduced by changing the product design, the process, or the control plan.

The purpose of the first stage of the algorithm is to quantify the magnitude of the selected problem. To do this, we monitor the output of the process using an appropriate sampling scheme (often a multivari plan) for a sufficiently long period of time, so that we see the effect of all large causes of variation, especially the dominant cause. The process variation is then displayed using a histogram or summarized numerically. This baseline histogram is called the Green $Y^{®}$ distribution (Shainin et al., 1997) in SS terminology.

We use the baseline distribution to quantify the problem, to set a goal that has the potential to improve the process, and to assess any proposed remedy. The baseline distribution is also used to plan and check that a dominant cause exhibits its full effect in each investigation in the progressive search. We call this the full range of variation. This is important information necessary to keep us from focusing on the wrong family of causes. The idea of quantifying the nature of the problem is part of all problem-solving approaches. The unusual feature of SS is the explicit link between the search for the dominant cause and the baseline distribution.

The second stage in the SS algorithm (see Figure 1) involves the quantification and establishment of an effective measurement system. Without a good measurement system, it is difficult to learn about and improve the process, and the measurement system itself may be home to the dominant cause of the problem. Having a separate step in the SS approach devoted to checking the measurement system helps to ensure this essential task is not neglected. We look at the recommended plan and analysis for assessing the measurement system in the next section.

By eliminating families of causes, SS reduces the number of specific inputs that are candidates for study. SS emphasizes checking the measurement system for the process output, but says little about establishing reliable measurement systems for any measured inputs.

The goal of the third stage of the SS algorithm is to generate clues about the dominant cause. This is the progressive search. At this stage, another key emphasis in SS is to "talk to the parts" (Shainin, 1992). In statistical jargon, we use observational rather than experimental plans as much as possible.

SS makes heavy use of observational plans such as multivari investigations, stratification, group comparison, and scatter (correlation) plots within the progressive search. It is surprising, given the availability of statistical software, that analysis of variance and regression techniques are not included. Recommended experimental plans, such as swapping components within assemblies are performed off-line and avoid disrupting production. The use of a sequence of observational plans is made explicit and is emphasized in SS in the search for the dominant cause unlike any version of Six Sigma we have seen.

The purpose of the fourth and the fifth stages of the algorithm is to confirm the identity of the dominant cause. The end result of the progressive search may be a single cause or a short list of suspects. With SS, dominant causes are verified using a formal experiment because of concerns about possible confounding (because of the earlier use of observational plans) and spurious associations (because of the small sample sizes). The suspect dominant causes are the factors that must be held fixed in the experiment. SS uses two level designs with the levels set at the ends of the normal range of variation of the suspect cause(s). Changing the levels of a dominant cause in the experiment should produce the full range of the output variation. Full factorial designs are recommended so that interactions among the suspects can be identified. With a single suspect, SS recommends a six run experiment (sometimes called B vs. C^{TM}-see the next section) with three replicates for each level.

A full factorial verification experiment is feasible because the list of suspects is short. Also, because the purpose is clear, there is little temptation to mix up the verification of the dominant cause and the search for a remedy. That is, at this stage, inputs that are normally fixed are not changed within the experiment. Note that these fixed inputs cannot be a dominant cause of the observed output variation.

We now discuss the steps of the algorithm in the remedial journey. We assume that a dominant cause

has been identified and verified. The first step in the remedial journey applies to the special case of a single dominant cause that is an interaction between two varying inputs. That is, a major component of the variation in the output, denoted by y, can be explained by the joint variation of two varying inputs denoted by x_1, and x_2. We have an interaction since the relationship between the output and the first input depends on the level of the second input. The presence of interaction suggests a non-linear relationship $y = f(x_1, \ x_2) + residual$, where the residual variation is relatively small, since x_1, and x_2 together are a dominant cause. We may be able to exploit this relationship to desensitize the process to variation in x_1 and x_2 by changing the set points of x_1, x_2 or both. We can investigate this possibility with a series of small experiments with two factors x_1, and x_2. This strategy may or may not be effective. It is a special application of parameter design where the experimental factors are limited to the set points of the inputs that make up the dominant cause. We can find little evidence (e.g., it is not explicit in the algorithm as shown in Figure 1) that SS considers the more general strategy to reduce variation due to an identified cause by exploiting interactions with a wider selection of inputs fixed in regular production. One of the referees pointed us to the Concept diagram, another SS tool. The only reference that we can find (Moore and Butkovitch, 1998) gives a Concept diagram that shows the effect of a process change on the output. We cannot see that this corresponds to applying parameter design and furthermore, there is no indication as to how to identify the necessary interactions between the normally fixed inputs and the varying dominant cause. One advantage of SS is that in the remedial journey, we deal only with dominant causes. This ensures that, if parameter design is used, the noise factors (dominant causes) are known to have a large effect. Resources are not wasted investigating unimportant noise factors. We can establish these specifications based on the specified tolerance of the output by quantifying the relationship between the output and the dominant cause. In SS, ^{TM} (Shainin, 1993b). This tool

is described in more detail in the Shainin Tools section.

The algorithm splits at the next stage. We take irreversible corrective action to mean that the variation in the cause can be eliminated. More interestingly, if this is not possible, then the algorithm suggests Process Control. In SS terminology, this means precontrol, not Shewhart control charts. Precontrol is a feedback control system applied to the dominant cause to keep its value within the specification limits derived in the previous stage. The relative merits of precontrol versus Shewhart or other control charts have been widely discussed—see the next section. Precontrol is a feedback controller designed as part of an adjustment scheme, and hence it should be compared to other feedback controllers, not just control charting. Feedback control can be effective only if the dominant cause exhibits structural variation (Joiner, 1994). That is, the dominant cause must vary in such a manner that we can predict the future from the current and past values, and then have time to make adjustments, as necessary. In SS language, if the dominant cause resides in the part-to-part family, no form of feedback control can be effective in reducing variation. If this is the case, then precontrol will not be effective and the algorithm provides no guidance as to how to proceed.

The final two stages of the algorithm need no further discussion.

We can find no evidence in the literature that strategies such as feedforward control, robustness and process desensitization are considered (Steiner and MacKay, 2005). If the dominant cause does not exhibit structural variation, then precontrol will fail as a process adjustment scheme.

The use of designed experiments on existing processes is common in all major industrial problem-solving approaches. However, in comparison to other approaches, in SS, the use of experimentation is subordinated to observational investigations. In particular, the Shainin system avoids screening experiments to look for a dominant cause. As described earlier, experiments are recommended in the diagnostic journey only after the list of suspect dominant causes is short. This is a major advantage of SS since observational investigations are typically

much cheaper and more easily implemented than experimental investigations. Since the dominant cause is acting in the current process, the view is that we can generate clues about its identity by watching the process in action in an informed organize way. We may not select the dominant cause as a factor in the experiment and it is often difficult to hold suspect dominant causes (that vary in normal production) fixed in an experiment, We need to choose extreme levels (after first determining what this means) of the suspect causes if we hope to establish that the cause is dominant. In our view the use of screening experiments to search for a dominant cause should be considered a tool of last resort.

SS is weak in its use of experimental plans in the remedial journey. To reduce variation, there must be a change in process settings, the control plan, or the product or process design; that is, a change to one or more process inputs that are fixed in normal production. These changes can be sought and investigated only by using experiments. In the remedial journey, screening experiments and the sequential approach to experimentation should be considered.

The stress on the importance of the measurement system is a strong point of the SS algorithm, shared by most versions of Six Sigma. This discussion of the importance of measurement systems is missing or limited, in many well-respected books on Statistical Process Control (SPC) and Design of Experiments (DOE), such as Montgomery (1996, 2001), and Ryan (1989).

The use of a systematic approach to problem solving is not unique to SS. There are many competitors such as DMAIC (define, measure, analysis, improve, control) in Six Sigma (Harry and Schroeder, 2000) See also Juran's (1988) Diagnostic and Remedial Journeys Approach, Harry's twelve-step Breakthrough Cookbook approach (1997, pp. 21.19), and the Process Improvement and Problem Solving Strategies proposed by Hoerl and Snee (2001). DMAIC maps well to the Shainin System, with D, M and A corresponding to the diagnostic journey and I and C corresponding to the remedial journey. Compared to these other systems, the purpose and the strategies for the individual stages in SS are more specific. The methodology is prescriptive, going as far as to

suggest specific tools that are useful for the different steps. The algorithm is specially designed for a medium to high volume manufacturing process in which inputs and outputs can be readily measured. Six Sigma, for instance, has a much broader range of application.

Another major difference is that SS does not distinguish between common and special causes as discussed earlier. Hoerl and Snee (2002), for example, suggest different systematic approaches to deal with common and special causes. SS, properly in our view, focuses on dominant causes.

A SELECTION OF SHAININ TOOLS

In this section, we describe and critique a selection of the more interesting and controversial tools associated with the Shainin System, namely: By tool, we mean both the plan of the investigation and the recommended analysis method. See Bhote and Bhote (2000) for a more extensive, though not complete, list of SS tools.

SS tools are generally statistically simple plans with small sample sizes that make extensive use of graphical displays and non-parametric tests that can be performed by hand. Given their purpose, we feel that the simple plans are to be highly recommended in most cases. We believe, however, that the non-parametric analysis methods are weak and non-intuitive. While we are strongly in favor of graphical approaches, with today's widespread availability of statistical software, ease of calculation is not an issue and we recommend supplementing the graphs with straightforward standard analyses. For some of the SS tools, we suggest alternative analysis methods that are better in most circumstances.

ISOPLOT®

An Isoplot® study (Traver, 1995; Shainin, 1992) is used to compare the relative size of the process and measurement system families of variation. In its simplest form, 30 units are selected, and each unit is measured twice. An Isoplot analysis starts with a scatterplot of the two measurements on each unit. On this plot, the horizontal variation is the overall process variation as measured by the first reading and
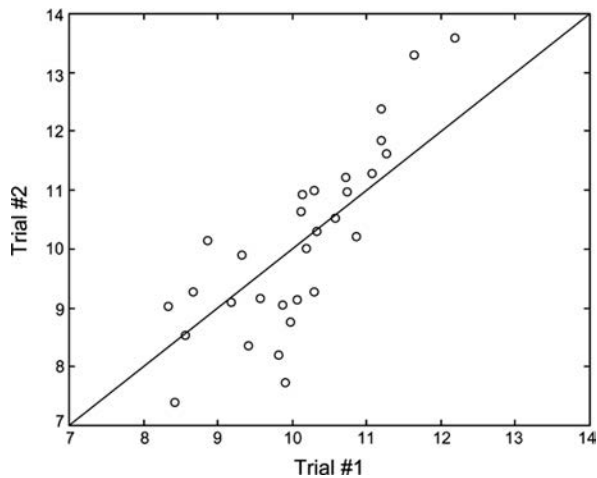
FIGURE 2 Scatterplot of measurement results.

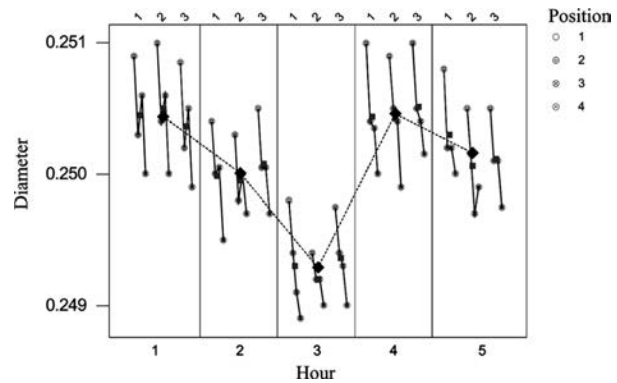**FIGURE 3** Multivari chart for diameter by position, part and hour.

the vertical variation is the overall process variation as measured by the second reading. The variation in a direction perpendicular to the 45-degree line represents the measurement system variation and, if all points lie near the 45-degree line, the measurement system variation is small. Figure 2 provides an example where, while not dominant, the variation due to the measurement system is relatively large.

With appropriately chosen pairs of measurements, we can assess repeatability or systematic differences between two operators, gauges etc. Outliers are obvious from the plot.

The SS Isoplot analysis includes specific rules for drawing an oval over the plotted points that can be used to numerically estimate the ratio of process to measurement variation, called the discrimination ratio. While plotting the data is a good idea, an analysis of variance (AIAG, 1995) is the preferred standard way to estimate the two variance components.

MULTIVARI

In a multivari investigation, we systematically sample from the process to capture the effect of various time and location based families of variation. Seder (1950a, 1950b, 1990) proposed a multivari chart to display such data. See also Snee (2001). A multivari is an excellent tool early in the progressive search for a dominant cause. It can be used at the beginning of the project to determine the Green Y distribution and simultaneously look for clues. Figure 3 shows a multivari chart using the diameter of a shaft as

the output. The shaft diameters are measured at four locations (left and right sides at different two orientations) for three shafts produced consecutively each hour. In Figure 3, we see there is little variation from shaft to shaft within an hour, some variation within shafts, and substantial variation from time-to-time, suggesting that the dominant cause must be an input that varies slowly, that is, that acts in the time-to-time family. This conclusion may be incorrect if we have not see most of the fall range of diameter variation established in the baseline investigation (i.e., the Green Y distribution).

A multivari chart provides a visual display of the components of variation associated with each family. However, when there is no obvious dominant family, it is useful to augment the plot with an appropriate analysis of variance to numerically estimate the variance components due to each family (see De Mast et al., 2001).

COMPONENT SEARCH[TM] AND VARIABLE SEARCH[TM]

Component Search (Shainin and Shainin, 1988) is used when units can be disassembled and reassembled without damage or change to any of the components or subassemblies. For ease of discussion, we do not distinguish between subassemblies and components. The goal is to compare the families of variation defined by the assembly operation and individual components. We start with two units, one "best of the best" (BOB) and one "worst of the worst" (WOW) with output values at the two extremes of the Green Y distribution. That is, we use leveraging

Shainin System[TM] for Quality Improvement

to ensure that we have the fall range of output variation. We eliminate families of causes by disassembling and reassembling, and possibly swapping components between the WOW and BOB parts. Applying component search, we first partition causes into two groups, the assembly and components families. If the assembly family can be eliminated (i.e., if repeated disassembly and reassembly of the BOB and WOW yield consistent results), the remaining causes are further subdivided into families defined according to individual components. There is a detailed four-stage investigation (confirmation, elimination, capping, and analysis, see Bhote and Bhote, 2000) to sort out which component family (or in the case of interactions, sets of families of components) is the home of the dominant cause. Component search is an experimental plan because we deliberately manipulate the inputs, i.e., the components. However it is performed off-line to avoid disruption of the production process.

We give an illustration of the results of a component search with four components in Figure 4. On the plot, the *Xs* correspond to the results for Assembly 1 and the *Ys* to the results for Assembly 2. Two units with extreme initial output values, given by the two left-most plotted points, were chosen to guarantee the full range of variation in the output. Then, the team disassembled and reassembled each unit two times. Since little change was observed in the output values in either the BOB or WOW, the results suggest that the dominant cause acts in the components family and not in the assembly family of causes. The dashed lines in Figure 4 give the performance averages for the first three output values. Next, by swapping components between the two
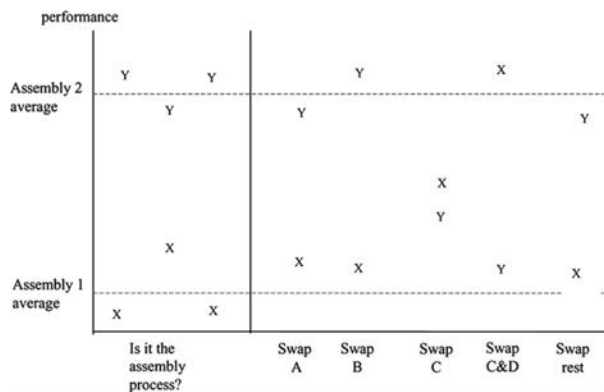

FIGURE 4   Component swap results.

assemblies one at a time and then pairwise, the dominant cause was identified as an interaction between varying inputs in components C and D.

The graphical analysis is effective when there is a dominant cause. However, since the order in which components are swapped is under the control of the investigators, the length of the search depends on their judgment regarding which component family is the likely home to the dominant cause. Amster and Tsui (1993) provide somewhat extreme artificial examples where component search yields incorrect conclusions. An alternative to the component swapping stage of component search is a 2^k factorial or 2^{k-p} fractional factorial experiment using the components as factors with levels defined by the WOW and BOB assemblies. An even more efficient process for eliminating component families, when feasible, is to proceed sequentially; that is, at each stage divide the remaining suspect components into only two subassemblies and swap one of the subassemblies—see Steiner and MacKay (2005).

In the first stage of Component Search, we must be careful that the off-line assembly/disassembly process matches normal production. Otherwise we may come to incorrect conclusions about the impact of the assembly family of causes.

Variable Search is similar to the component swapping stages in Component Search. It is used to identify a dominant cause, when the progressive search produces a list of four or more suspects, and no other simple investigation can rule out any of these possible dominant causes. With three or fewer suspects, SS recommends a full factorial experiment to identify the dominant cause. In Variable Search, the first steps are to list the suspects in order of expected importance and to determine two levels for each, based on their range of variation in normal production. Next, through trial and error, the two levels of each input are assigned labels "high" and "low" so that the two runs with all inputs at the same level (all high or all low) produce output levels that are at the extreme ends of the Green Y distribution. Finally, the levels of each suspect are varied one at a time or pair wise as in component search to find the dominant cause.

Variable Search is an online experiment, with all of the difficulties of setting or holding the varying inputs at their extreme levels. The ordering of the suspects and the determination of their levels can

be difficult and take substantial time and effort. The experiment cannot be successful if these levels are not correctly determined. It may be difficult to assign the high and low labels, especially if the dominant cause is an interaction. The length of the search depends on how well the suspects are ordered and the complexity of the dominant cause. See Ledolter and Swersey (1997b) for a critical view of variable search. We agree with their conclusion that fractional factorial designs are generally a better approach than Variable Search. This situation in which there is a long list of specific suspects seems ideally suited to an observational plan that uses multiple regression, a tool that does not appear to be part of SS, to reduce the number of possible dominant causes.

GROUP COMPARISON

Group comparison has two uses. If the problem is defined by a binary output (such as defective or not), we can use group comparison to try to identify a continuous output to reformulate the problem. This is especially useful if the defect is rare where group comparison is akin to a case control study. We can also use group comparison to identify specific suspect causes late in the progressive search after other investigations have eliminated many large families of causes of variation from consideration.

With group comparison (Bhote and Bhote, 2000), we select two groups of three or more parts with different values of the binary output or with extreme values of a continuous output. This is another application of leveraging. We measure the parts on as many input characteristics as possible, consistent with previously generated clues. If a measured input is a dominant cause, the values of this input will be systematically different between the two groups of parts.

The recommended analysis for each measured input is a two sample nonparametric test that requires either complete separation of the BOBs and WOWs or a minimum "endcount" (Bhote and Bhote, 2000) to identify a suspect dominant cause. Endcount is due to Tukey (1959) who dubbed the test "compact" because the test statistic can be calculated easily; the critical values are essentially independent of sample size and can be carried in the analyst's head. We suggest a standard analysis based on plots and t-tests. If there is a large effect (i.e., one of the inputs measured is a dominant cause), we can find the cause using only the plots of the data. Two way interactions can be seen by looking at all scatterplots of the suspect causes with different plotting symbols for the BOBs and WOWs. Since the comparisons are usually based on small sample sizes, there is a risk of confounding and also a strong possibility of identifying spurious causes because of the multiple testing.

Figure 5 illustrates the typical analysis. The data arose from a group comparison to help find a dominant cause of leaks in the rear intake wall of engine blocks. The output was binary; there was no measure of the size of the leak. Whenever the team found an intake wall leaker, they also set aside a non-leaking block. They collected 50 leaking and 50 nonleaking blocks. Then, for each of the sampled blocks, they measured thickness (in inches) at six locations in the left rear intake wall. To analyze the data, we construct side-by-side boxplots of wall thickness at each location for leakers and
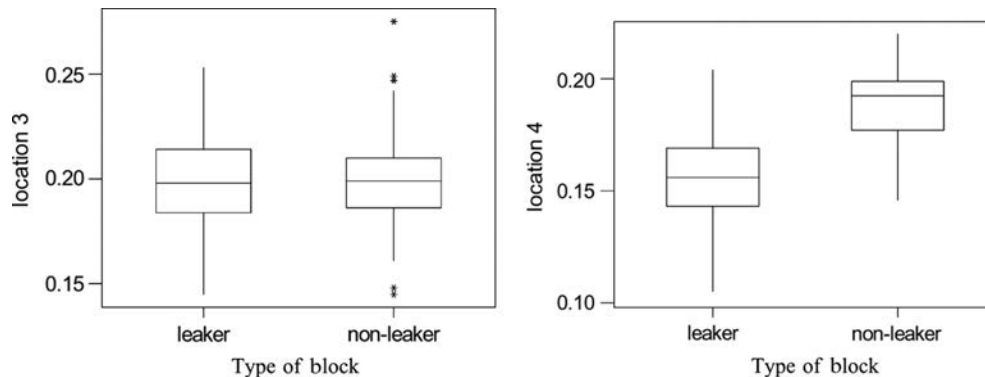

FIGURE 5   Boxplots of locations 3 and 4 wall thickness by block type.

non-leakers. We show the results for two locations in Figure 5. The right-hand plot shows a clear difference in median wall thickness between leakers and non-leakers at location four. There was little difference for the other locations as in the left-hand plot. The team concluded that wall thickness at location four was a dominant cause of rear intake wall leaks.

A version of group comparison called Paired Comparisons[TM] (Shainin, 1993b, and Bhote and Bhote, 2000) involves pairing or matching of the defective and non-defective units. In the proposed analysis, the BOBs and WOWs are paired, usually based on time of production. Shainin (1993b) writes "Paired Comparisons are appropriate when the largest family of variation is part to part." In this context, since we are looking for a dominant cause, pairing adds to the complexity of the plan and little value. In statistical experiments we use pairing to eliminate the risk of confounding and to increase the precision of the conclusions about the experimental factor in the presence of other varying inputs that have a large impact on the output. If the dominant cause acts in the part-to-part family, paired comparisons will produce pairs that are similar only with respect to other inputs that have little influence. Thus, unless a Pink X (a second large cause) acts time to time, pairing will decrease the precision of the conclusions. This loss may be important due the recommended small sample sizes.

A paired comparison conducted on arbitrary constructed pairs has been suggested in Bhote and Bhote (2000). With arbitrary pairs, the conclusions of the analysis depend on the way pairs are produced and, on average, the sensitivity of the procedure will be lower than that of the unpaired analysis. In general, pairing seems a bad idea in this context.

TOLERANCE PARALLELOGRAM[TM]

A tolerance parallelogram is used to establish appropriate specification limits for a dominant cause. We select a number of parts with output values that cover the full range of variation and measure the value of the output and the dominant cause on each selected part. Constructing a scatterplot and, using a specified proprietary procedure, we derive the tolerance limits for the dominant cause from the output specifications taking into account the residual variation in the output. See Figure 6 where we used
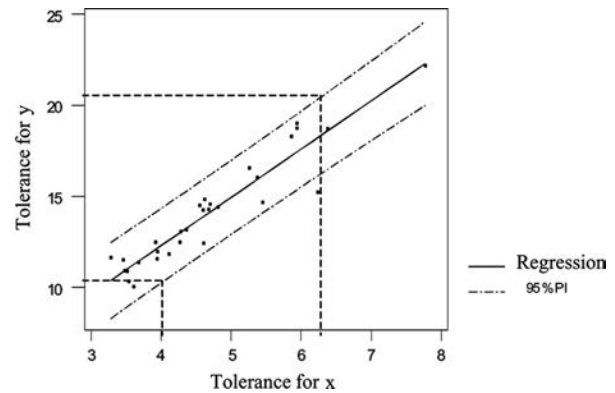
FIGURE 6 Using regression to set specification limits (tolerances) for the dominant cause (x) given a tolerance for the output (y).

prediction intervals from a simple regression model for this task. The idea is that if we control the dominant cause within its derived tolerance range, the output will be controlled with the desired specifications. At this stage, there is no effort to determine how to exert the required control of the dominant cause. If the residual variation is too high (e.g., the cause is not sufficiently dominant), then there will be no tolerance left for the cause. We can extend the methodology to the cases where a dominant cause is an interaction between two inputs or where there is more than one dominant cause, using a more complex model.

B vs. C[TM] AND FACTORIAL EXPERIMENTS

B vs. C is a simple experimental plan used to compare two treatments or process conditions represented by the letters B and C. One use in SS is to verify that an identified cause is dominant after other clue generation tools have led to a single suspect. A second use is to validate a solution when, for example, the goal is to shift the process center or reduce a defect rate. In the validation application, the letters B and C denoted the "better" (we hope) and "current" conditions. Note that for verification of a suspect cause, the better and current terminology may not be appropriate.

In the simplest recommended plan, three units are produced under treatment B and three under treatment C. Bhote and Bhote (2000) call this the "six-pack test." The levels for the suspect dominant cause for the B and C runs are selected at the

extremes of the variation of the suspect in normal production. The order of the six runs is randomized. The recommended analysis is based on the end count scheme discussed in the group comparison section. Only if the output values for the three B runs and the three C runs separate in the expected direction have we verified the dominant cause. Tukey created this test as a one sided test of hypothesis; no change versus change is a specified direction. A sample size of six units has low power but, by taking larger samples, power can be increased. Since we must see most of the full range of variation in the output if we have a dominant cause, the formal hypothesis test is essentially irrelevant here.

When validating a solution, the use of the compact end count test is undesirable since the loss of power versus a wide selection of parametric or other nonparametric tests could lead to the abandonment of an improved way of operating the process.

SS makes use of full factorial experiments to isolate a dominant cause among a short list of suspects (Shainin and Shainin, 1990). The plan and implementation of the experiment with its careful attention to the selection of levels of the suspects and the use of randomization is highly recommended, as is the use of plots of the data. Here the formal analysis based on a sequence of end-count tests leaves much to be desired. The first step is to calculate the effects, and then examine the significance of the largest, ignoring the selection effect, by rank ordering the output based on the levels of the selected factor. Next, the second largest effect is formally tested by rank ordering and determining the end count of the residuals from the first analysis. In this way, we have removed the effect of the Red X. And so on for the smaller effects. This procedure has the colorful name Pink X^{TM} shuffle (see Shainin and Shainin, 1990 for a detailed description). It is opaque and suffers from both selection effects and multiple testing issues. At each stage, the test is not based on the residual variation as established by the experiment, but also includes the variation due to the other factors being studied. This reduces the sensitivity of the method at the first step and can be devastating at the second step. To our knowledge, no one has extended the Tukey method to factorial experiments. We suggest a standard analysis using effect plots, probability plots of the effects and an analysis of variance to complement the excellent design.

PRECONTROL

Precontrol (also called stoplight control), first introduced by Satterthwaite (1954), is used to signal the need for a process adjustment. In SS, precontrol is applied to the dominant cause using specification limits developed with a Tolerance Parallelogram^{TM} as described above. Shainin (1995) writes "If the Red X can't be controlled with an irreversible corrective action, then precontrol needs to be put on the Red X. SPC [Precontrol] is always more effective when it is used on the Red X instead of the Green Y."

To implement precontrol, parts are sampled and measured according to a periodic schedule. The specification range is divided into three zones as illustrated in Figure 7:

- Green is go, and for a two sided tolerance occupies the middle half of the specification range,
- Yellow is the warning zone and covers the outside quarters of the specification range,
- Red is stop and includes anything outside the specification range.

Precontrol is conducted using the following rules (there are many variations on this theme):

i. Set-up or after an adjustment: OK to run when five parts in a row are green.
ii. Running: Sample and measure two consecutive parts on a fixed frequency.

- If first part is green, continue to run.
- If first is yellow, check the second part – if it is yellow or red, stop and adjust.
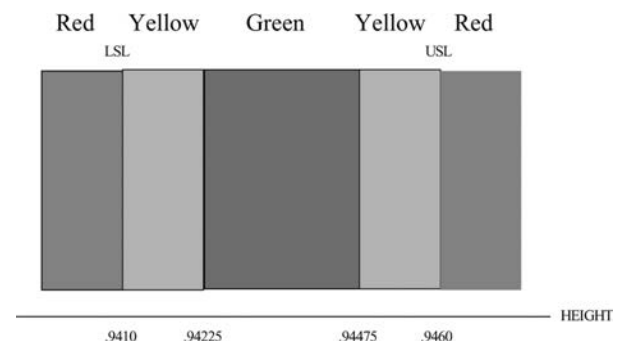- If first is red, stop and adjust process.

FIGURE 7 An example of precontrol zones.

Of all the tools prominent within SS, precontrol has received the most attention in the research literature. For example, see Shainin (1984), Traver (1985), Salvia (1987, 1988), Shainin and Shainin (1989), Mackertich (1990), Gruska and Heaphy (1991), Ermer and Roepke (1991), Ledolter and Swersey (1997a), Steiner (1997–1998).

To be successful, precontrol requires good specification limits and a process that operates within these limits in the short term. Otherwise, it will be difficult to get five parts in a row in the green zone to start. Since precontrol is a feedback adjustment scheme, it can only be effective if the process drifts slowly or jumps and sticks. Precontrol may result in increased variation if used on a process that has large part-to-part variation. Although it is often compared to statistical process control (SPC), the goal of precontrol is to identify the need for adjustment. It is not useful for process monitoring nor for the identification of the action of special causes. More sophisticated control and feedback schemes, such as proportional-integral-derivative (PID) controllers (see del Castillo, 2002), are alternatives that may yield better results. Note that, while precontrol signals the need for an adjustment, it does not include an adjustment rule which is required to implement the system in practice.

SUMMARY

The guiding principles of the Shainin System are powerful, and, at least in combination, unique. They include the application of Juran's Pareto principle to the contribution of the causes, the emphasis on using observational investigations in the diagnostic journey, the search for a dominant cause using the process of elimination and the use of leveraging. SS deals carefully with the problem of possible confounding of suspect causes by conducting a small verification experiment. We think that the principles and tools related to the diagnostic journey are generally very strong. Those related to the remedial journey are much weaker. This may be the case because once a dominant cause is identified, in some instances, the remedy is obvious and no further investigations are needed.

The Shainin System, as reflected by the genesis of the methodology in manufacturing, is best suited for medium to high volume production. Most of the tools implicitly assume that many parts are available for study. When using leverage, where the investigations involve only a small number of parts, there must be a substantial amount of measurement to find the parts with extreme values. Like many other systems with strong statistical components, SS does not handle well situations where there are few parts to "talk to" such as in the design and development of new products or processes.

Although our assessment of SS is strongly positive, there are some unfortunate aspects about its promotion. Most notably, many of the specific tools and the whole approach have not been subject to a peer reviewed public discussion. This may be because much of the specific terminology is trademarked and is thus legally protected. We feel this is unfortunate since it has reduced the dissemination of what we think is an excellent approach. Also, some books that promote the methodology, such as Bhote and Bhote (2000), are full of unhelpful hyperbole that limits discussion of feasible alternatives. In our experience, there is also a rigidity with which the methodology is presented. In many situations, other statistical tools, such as regression, time series, and analysis of variance, could be very useful, but are not employed because they are not formally part of the SS tool bag.

ACKNOWLEDGMENTS

ABOUT THE AUTHORS

Stefan Steiner and Jock MacKay are both Associate Professors in the Statistics and Actuarial Science Department of the University of Waterloo. They are also active consultants who have worked with organizations from a wide range of industries, including automotive, telecommunications, aerospace, government, and more.

John S. Ramberg, Professor Emeritus, University of Arizona and quality consultant, is a fellow of ASQ,

ASA and IIE. He holds a B Elec Engr (Ind Engr), Minnesota; MS OR, Cornell; PhD Engr Stat, Cornell. He served on the Iowa faculty (stat and mgt engr) from 1967-78, and the Arizona faculty, Sys & Ind Engr from 1978–2003, where he also was head. His work experience includes Procter & Gamble, Motorola, Hughes Space Systems. He is editor emeritus of The Journal of Quality Technology, and member of the editorial board.

REFERENCES

Amster, S., Tsui, K. L. (1993). Counterexamples for the component search procedure. *Quality Engineering*, 5:545–552.

Automotive Industry Action Group (AIAG). (1995). *Measurement Systems Analysis*, 2nd ed. AIAG Southfield, MI.

Bhote, K. R. (l988). *World Class Quality: Design of Experiments Made Easier, More Cost Effective Than SPC*. Saranac Lake, NY: American Management Association (AMACOM).

Bhote, K. R. (1991). *World-Class Quality: Using Design of Experiments to Make It Happen*. NY: American Management Association.

Bhote, K. R., Bhote, A. K. (2000). *World Class Quality*, 2nd ed. NY: American Management Association.

De Mast, J., Schippers, W. A. J., Does, R. J. M. M., Van den Heuvel, E. (2000). Steps and strategies in process improvement. *Quality and Reliability Engineering International*, 16:301–311.

De Mast, J., Roes, K. C. B., Does, R. J. M. M. (2001). The multi-vari chart: A systematic approach. *Quality Engineering*, 13:437–448.

Del Castillo, E. (2002). *Statistical Process Adjustment for Quality Control*. New York: Wiley and Sons.

Does, R. J. M. M., Roes, K. C. B., Trip, A. (1999). *Statistical Process Control in Industry*. Dordrecht, The Netherlands: Kluwer Academic Publisher.

Doyle, A. C. (1994). The Adventures of Sherlock Holmes: The adventure of the beryl coronet. Reissue edition (first published 1892). Berkley, CA: Berkley Publishing Group.

Ermer, D. S., Roepke, J. R. (1991). An analytical analysis of precontrol. American Society for quality control, Quality Congress Transaction—Milwaukee, WI, pp. 522–527.

Gruska, G. F., Heaphy, M. S. (1991). Stop Light Control—Revisited. American Society for Quality Control Statistics Division Newsletter, Fall.

Harry, M. J. (1997). *The Vision of Six Sigma: A Roadmap for Breakthrough*, 5th ed. Phoenix, AR: Tri Star Publishing.

Harry, M., Schoeder, R. (2000). *Six Sigma: The Breakthrough Strategy Revolutionizing the World's Top Corporations*. New York: Doubleday.

Hockman, K. (1994). Review of *World Class Quality*, by K. Bhote, *Technometrics*, 36:425–426.

Hoerl, R. W., Snee, R. D. (2003). *Statistical Thinking, Improving Business Performance*. Pacific Grove, CA: Duxbury.

Joiner, B. J. (1994). *Fourth Generation Management: The New Business Consciousness*, McGraw-Hill, New York.

Juran, J. M. (1988). *Juran's Quality Control Handbook*, 4th ed. New York: McGraw-Hill.

Juran, J. M., Gryna, F. M. (1980). *Quality Planning and Analysis*, 2nd ed. New York: McGraw-Hill.

Ledolter, J., Swersey, A. (1997a). An evaluation of precontrol. *Journal of Quality Technology*, 29:163–171.

Ledolter, J., Swersey, A. (1997b). Dorian Shainin's variables search procedure: A critical assessment. *Journal of Quality Technology*, 29:237–247.

Logothetis, N. (l990). A perspective on Shainin's approach to experimental design for quality Improvement. *Quality and Reliability Engineering International*, 6:195–202.

Mackertich, N. K. (1990). Precontrol vs. control charting: A critical comparison. *Quality Engineering*, 2:253–260.

Montgomery, D. C. (1996). *Introduction to Statistical Quality Control*, 3rd ed. New York: John Wiley and Sons.

Montgomery, D. C. (2001). *Design and Analysis of Experiments*, 5th ed. New York: John Wiley and Sons.

Moore, D. (1993). Review of *World Class Quality*, by K. Bhote, 1991. *Journal of Quality Technology*, 21:76–79.

Moore, D., Butkovitch, S. Identifying sources of solder defects in printed circuit assemblies incorporating ball grid array packages. http://www.smta.org/files/SMI98-MooreDavid.pdf.

Nelson, L. S. (1991). Review of *World Class Quality*, by K. Bhote, 1998. *Journal of Quality Technology*, 25:152–153.

Ryan, T. P. (1989). *Statistical Methods for Quality Improvement*. New York: John and Wiley and Sons.

Salvia, A. A. (1987). Performance of PRE-control sampling plans (Corr: Vol. 19 p. 244). *Journal of Quality Technology*, 19:85–89.

Salvia, A. A. (1988). Stoplight control. *Quality Progress*, September, pp. 39–42.

Satterthwaite, F. E. (1954). A simple, effective process control method. *Rath & Strong Inc*. Boston, MA: Report 54-1.

Seder, L. A. (1950a). Diagnosis with diagrams—Part I. *Industrial Quality Control*, January.

Seder, L. A. (1950b). Diagnosis with diagrams—Part II. *Industrial Quality Control*, February.

Seder, L. A. (1990). Diagnosis with diagrams. *Quality Engineering*, 2:505–530 (reprinted from original in *Industrial Quality Control*, 1950).

Shainin, D. (1984). Better than good old $\overline{X}$ and R charts asked by vendees. 40th Annual Quality Congress Proceedings, ASQC, pp. 302–307.

Shainin, D. (1986). Better than Taguchi orthogonal tables. 42th Annual Quality Congress Proceedings, ASQC, pp. 446–451.

Shainin, D., Shainin, P. (1988). Better than Taguchi orthogonal tables. *Quality and Reliability Engineering International*, 4:143–149.

Shainin, D., Shainin, P. (1989). Precontrol versus $\overline{X}$ & R charting: Continuous or immediate quality improvement? *Quality Engineering*, 1:419–429.

Shainin, D., Shainin, P. D. (1990). Analysis of experiments. 45th Annual Quality Congress Proceedings, ASQC, pp. 1071–1077.

Shainin, P. D. (1992). Managing SPC, a critical quality system element, 46th Annual Quality Congress Proceedings, ASQC, pp. 251–257.

Shainin, P. D. (1993). Managing quality improvement. 47th Annual Quality Congress Proceedings, ASQC, pp. 554–560.

Shainin, P. D., Shainin, R. D., Nelson, M. T. (1997). Managing statistical engineering. 51st Annual Quality Congress Proceedings, ASQC, pp. 818–532.

Shainin, R. D. (1992b). Technical problem solving strategies, a case study. 46th Annual Quality Congress Proceedings, ASQC, pp. 876–882.

Shainin, R. D. (1993b). Strategies for technical Problem solving. *Quality Engineering*, 5:433–438.

Shainin, R. D. (1995). A common sense approach to quality management. 49th Annual Quality Congress Proceedings, ASQC, pp. 1163–1169.

Snee, R. D. (2001). My process is too variable—now what do I do?: How to produce and use a successful multi-vari study. *Quality Progress*, December, 65–68.

Steiner, S. H. (1997–1998). Precontrol and some simple alternatives. *Quality Engineering*, 10:65–74.

Steiner, S. H., MacKay, R. J. (1997–1998). Strategies for variation reduction. *Quality Engineering*, 10:125–136.

Steiner, S. H., MacKay, R. J. (2005). *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Process*. Milwaukee: Quality Press.

Traver, R. W. (1985). Precontrol: A good alternative to $\overline{X}$-R charts. *Quality Progress*, September, pp. 11–14.

Traver, R. W. (1995). *Manufacturing Solutions for Consistent and Reliability*. New York: American Management Association.

Tukey, J. W. (1959). A quick, compact, two-sample test to Duckworth's specifications. *Technometrics*, 1:31–48.

Ziegel, E. R. (2001). A review of *World Class Quality*, 2nd ed. By K. Bhote, A. Bhote, *Technometrics*, 43:112–113.