Durham E-Theses

Concise Process Improvement Methods

COX, STEVEN

Concise Process Improvement Methods

Steven Cox BEng

MSc by Research

School of Engineering and Computing Sciences

University of Durham

2011

Steven Cox

Concise Process Improvement Methods Abstract

This thesis reviews two methodologies for process improvement; Six Sigma and the Shainin System. A strengthened methodology is developed following the 12-step Six Sigma DMAIC cycle with an added Shainin loop in the Analyse phase to narrow down sources of variation. This Hybrid Six Sigma framework is used to develop a sampling strategy known as the Process Variation Diagnostic Tool (PVDT).

The PVDT allows a Gage R&R and a Provisional Process Capability study to be carried out with just 20 samples. It also allows for an Isoplot^{SM} and a Shainin Multi-Vari study. The method was then reviewed in three different industrial situations to demonstrate its effectiveness. Applying the PVDT allowed the project teams involved to quickly produce Gage R&R and Provisional Process Capability Studies. It reduced samples required from the combined 110 measurements from 60 products typically taken in industry to 60 measurements from 20 products. A significant advantage was the ability to extract a Shainin Multi-Vari Study from measurements taken for the PVDT. This technique allowed the project team the ability to categorise the most significant families of variation. From these case studies it can be seen that at the border of the Measure/Analyse phase in Six Sigma the proposed PVDT offers an efficient method of collecting Six Sigma metrics and steering the course of an improvement project.

A teaching vehicle known as the PIM game is introduced to demonstrate and facilitate the teaching of a number Process improvement Method. These methods are directly related to Six Sigma and Shainin methods developed in this thesis. The historical development and need for a teaching game are discussed.

Finally the thesis proposes a new method of destructive measurement system analysis (MSA). An industrial problem is used to benchmark the method against a traditional approach to destructive MSA. The project highlights when there is a second non-destructive test a conservative estimate of Gage R&R can be determined for destructive test equipment.

Table of Contents

List of Figures

List of Tables

Nomenclature

ANOVA	Analysis of Variance
AV	Appraiser Variation
BOB	Best of the Best
C_p, C_{pk}, P_p and P_{pk}	Process Capability Metrics
CTQ	Critical-to-Quality
DMAIC	Define, Measure, Analyse, Improve and Control
DOE	Design of Experiments
DPMO	Defects per Million Opportunities
EV	Equipment Variation
FET	Field Effect Transistor
LSL	Lower Specification Limit
PIM	Process Improvement Method
PVDT	Process Variation Diagnostic Tool
R&R	Repeatability and Reproducibility
SS	Shainin System
USL	Upper Specification Limit
WOW	Worst of the Worst

Acknowledgements

I would like to thank my supervisors Prof. Val Vitanov, for advice on my work over the previous 18 months and John Garside, for the guidance and direction behind this project.

I'm also grateful to Mike O'Neill, Steve Endacott and Lee Costello at TRW for hosting live industrial projects to demonstrate some of the theory developed, to Adam Bent and Steve Richardson for setting up test equipment for the Durham end of the project, Anthony Middleton for assisting in testing and the Durham Graduate School for providing the funding for the TRW project.

I would finally like to thank Lucy Eland for her constant support over the previous year and for her assistance compiling and binding this report.

1   Introduction

The purpose of this thesis is to introduce and examine two process improvement methodologies in use in industry; Six Sigma and the Shainin System. The importance of scrutinising Six Sigma and Shainin is identified by Aboelmaged [1] and Senapti [2]. Aboelmaged [1] noted from the literature on Six Sigma the need for academics to develop a deeper and richer knowledge of Six Sigma so they do not over hype or quickly dismiss it. Whilst Senapti [2] points out that there has been little discussion or exposure of the Shainin System.

This thesis will explore many techniques used by Six Sigma and Shainin System to gain a full understanding of them and their merits. The result of this analysis will provide a platform to develop process improvement methods and methodologies to expand their use. Case studies and industrial projects are used to test the implementation of new methods in real industrial situations.

A teaching vehicle known as the PIM game is introduced to demonstrate and facilitate the teaching of a number Process improvement Method. These methods are directly related to Six Sigma and Shainin methods developed in this thesis. The historical development and need for such a teaching game are discussed. Trial developments introduced by Cox are fully tested and further developments to the game are explored.

2   Process Improvement Methodologies

This chapter will explore the literature of two process improvement methodologies. The methodologies explored will be the Six Sigma methodology for process improvement and the Shainin System (SS). Similarities and differences, strengths and weaknesses will then be drawn between the two methodologies to derive a strengthened methodology. This will provide context for the techniques and tools developed during the remainder of the thesis.

2.1   Six Sigma

Development of Six Sigma methodology for process improvement was started by Motorola in 1979, at a time when most American companies believed quality cost money [3]. The original approach was for use in manufacturing to improve quality and reduce costs, contrary to popular belief at the time that high quality costs. The company's emphasis focused on a number of advanced quality tools to prevent defects and achieve bottom line results. In the early 1990's the methodology soon spread to other American based companies including General Electric, Allied Signal, and Texas Instruments [3][4].

In process improvement the goal of a Six Sigma project is to reduce the number of defects a process produces and the variation present in Critical-to-Quality characteristics (CTQ), which are determined by customer requirements, to 3.4 Defects per Million Opportunities (DPMO) or a yield of 99.9997%. A quality level of 3.4 DPMO correlates to the 6σ level deemed by many texts as world class for a manufacturing process [3][4]. Six Sigma sets out to achieve this by employing a structured and systematic methodology known as DMAIC (Define, Measure, Analyse, Improve and Control), see Figure 1. As its name suggests it heavily uses statistical tools within the DMAIC framework to achieve data driven improvement to customer driven CTQs.

In the extensive review of Six Sigma literature by Aboelmaged [1], it was identified that there were differences in Six Sigma definitions depending on the context it was applied in. Although the original implementation at Motorola was process improvement in a manufacturing setting, its successes soon led to the principles being applied outside the manufacturing division to business strategy [3]. This difference in applications, is described by De Mast et al. [5], as the inner-MAIC loop (MAIC referring to the core of the DMAIC strategy) where Six Sigma is a process improvement strategy and the Outer-MAIC loop where its application to wider business strategy is acknowledged. As the remainder of the text focuses on process improvement in manufacture, it will use the definition of the inner-MAIC loop.
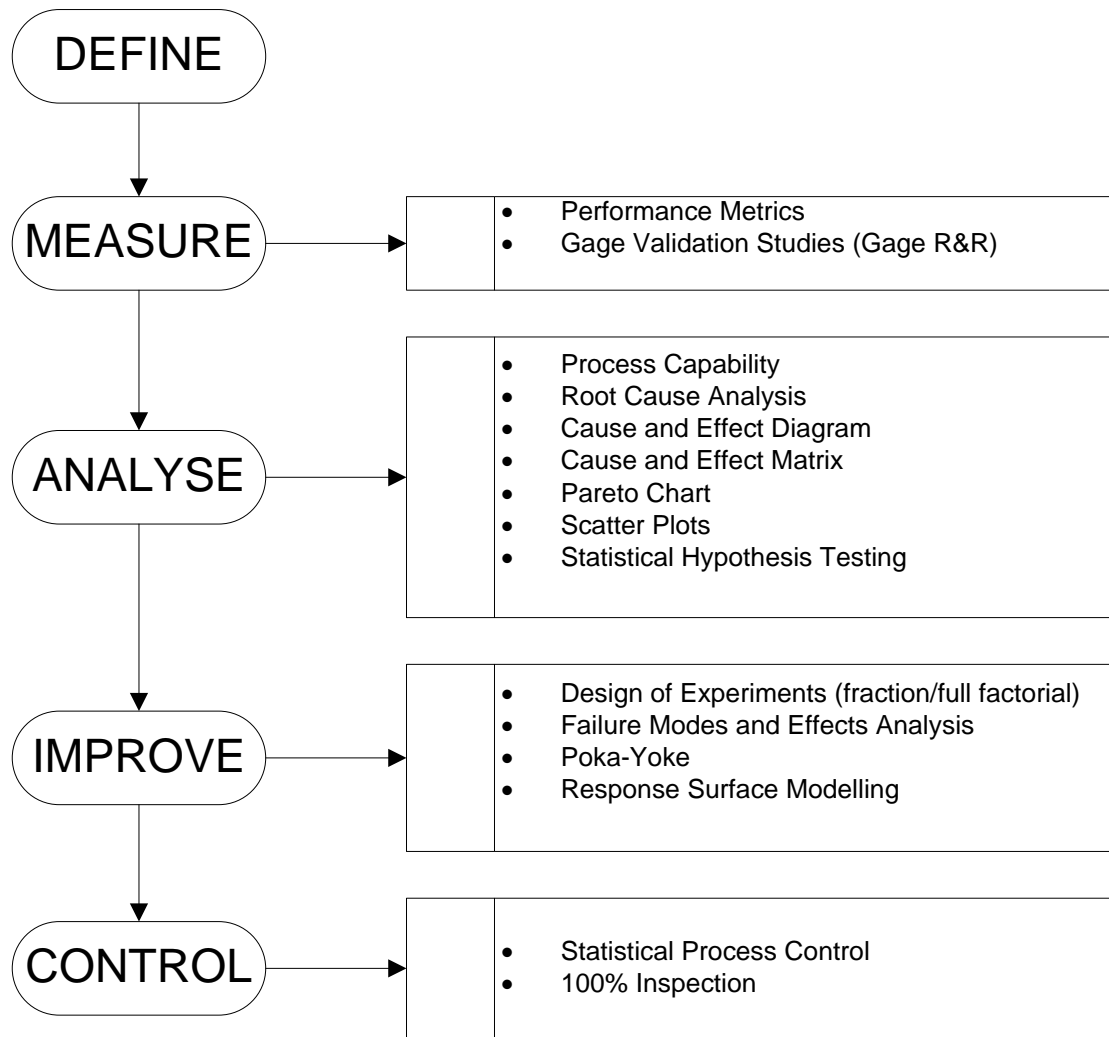
Figure 1 DMAIC Six Sigma Improvement Model with Tools, Adapted from [4]

As previously stated, when a Six Sigma quality improvement project is undertaken it is common to follow the five phase improvement cycle DMAIC [4][6], see Figure 1. In the Define phase, the problem is captured; potential benefits of the project are assessed. In the Measure phase, measurement capability is established and current performance levels are determined. In the Analyze phase, root causes of defects are uncovered. In the Improve phase, the influences of key process variables are quantified and the process modified to reduce defect levels. In the Control phase, actions are taken to sustain the improved level of performance.

Six Sigma texts, like Pande et al. [4] and George et al. [6], outline many techniques and tools that can be used at each stage of the quality improvement cycle. However when it comes to the Analyse phase they often jump from an extremely subjective approach, using brainstorming and cause-and-effect diagrams to form casual hypothesis, to complex statistical tools to validate these casual hypothesis. This weakness in Six Sigma's "exploration" is pointed out in De Mast [7] and it is the purpose of this research to

introduce techniques to the methodology to improve this shortcoming with the traditional Six Sigma approach.

This weakness is particularly important to overcome when the cost of sampling is very high or a low volume of product is available to test. In this situation extremely complex Designs of Experiments (DOE) can be impractical, for example applying a 2-level Full Factorial design when there are 5 or more factors is cumbersome (with 5 factors a minimum of 2^5 or 32 experiments is needed). Using less powerful screening techniques such as Fractional Factorials or Plackett-Burnam will reduce the numbers of experiments needed but at the expense of higher order interaction effects and the numbers of experiments needed can still spiral out of control if there are large numbers of factors present (with 15 factors 32 experiments are needed for a Fractional Factorial design). Other approaches used in the Six Sigma methodology to identify important factors are scatter plots and cause-and-effect matrices. Scatter plots are a method of finding correlations between factors graphically, but they can lead to potentially erroneous results as correlations can appear as a result of coincidence or two factors examined are linked by a related underlying cause [6]. Cause-and-effect matrices offer a method of linking input factors to outputs, but are extremely subjective. Importantly the real root cause of a quality problem could be missed if DOE is applied based on casual hypothesis techniques, such as cause-and-effect matrices, to identify important factors as root causes of variation may be eliminated through subjective hypothesis.

It is the authors belief that although this overall strategy is strong, the techniques used to implement the measure and analyse phases of the improvement cycle are weak. This is in line with De Mast [7] analysis of quality improvement methodologies that *"Six Sigma seems the most complete strategy"* but *"the guidance and tools that are given for the exploration phase lack clear structure and coherence"*.

2.2 The Shainin System[SM]

Originally Shainin strategies, known as Statistical Engineering, where primarily developed by Dorian Shainin beginning with the Lot Plot in 1943 [8]. Throughout his consultancy career he continued developing convergence techniques and the Shainin System[SM] (SS) ([SM] is a service mark of Shainin Consultants Inc.), a problem-solving strategy which is outlined in Shainin [9] and Steiner [10], see Figure 2.

There has been little peer review work of Dorian Shainin methods. Nor have they been exposed at large to professionals because of proprietary reasons [2]. There is a description

of some Shainin techniques in Steiner [10] but the most complete descriptions are in Bhote [11][12].

Define the Project

Establish Effective Measuring System → • Isoplot

Generate Clues → • Multi-Vari • Components Search • BOB v WOW

List Suspect Variables

Statistically Designed Experiment → • Variables Search • Full Factorial

Red X Found?

No

Yes

Interaction? —Yes→ Optimise → • EVOP • REVOP

No

Realistic Tolerances

• Tolerance Parallelogram

Irreversible Corrective Action?

No

Yes

Statistical Process Control → • Pre-Control

Monitor Results

Customer Enthusiasm!

Figure 2 the Shainin System^{SM} with Tools from [9]

These texts have been heavily criticised by Hockman [13] and Ziegel [14] for being self promotional and for their dismissal of classical DOE techniques. A detailed critique of Bhote first edition [11] is in Hockman [13], stating *"the book jacket and forward contain primarily hype. Heavy use of advertising phraseology is continued throughout the text."* and concluding *"the Shainin techniques have some interesting twists.... I would be very*

concerned, however, if the reader abandoned correct application of sound (classical) DOE tools as recommended." A briefer, but more scathing, review of Bhote second edition [12] given by Ziegel [14] concluding *"use this book at your own risk. It is more than a little extreme."* Although the books outline many useful Shainin techniques, they do them a *"disservice since the hyperbole hides many of the genuinely useful ideas"* [10].

Shainin [9] states *"the quality professionals real job is to assure that we always produce product or service within specified tolerances."* This is achieved by finding the major and secondary causes of variation, known in SS as the Red X^{SM} and Pink X^{SM} respectively, which affect the output, known as the Green Y^{SM}, in a problem process. To determine the Red X^{SM} the Shainin Convergence techniques and DOE are implemented within the SS algorithm (see Figure 2). Corrective action and/or statistical process control (SPC) are finally implemented to control the Red X^{SM}. SS establishes a *"Generate Clue"* phase which uses offline techniques to eliminate variables in a process which do not have an effect on the overall variation without disrupting process settings. This allows DOE to be performed with the identified suspect variables with fewer experiments needed to find the root cause or Red X^{SM}, this approach limits the online testing which causes disruption to a process. Steiner [10] supports this view stating *"the [SS] algorithm is very strong for the diagnostic journey."*

2.3   Comparison

The SS algorithm in Figure 2 shows us there are many similarities between the overall Six Sigma and SS methodologies for process improvement. They both use rhetoric which is comparable, when referring to measured outputs SS talks about the Green Y^{SM} where as Six Sigma refers to CTQ's. When analysing inputs which affect a process SS labels important factors as the Red X or Pink X and Six Sigma terms these factors as the root causes of variation. The difference in terms shows that Six Sigma and SS use a different language to one another but they stem from a difference in their historical development rather than a difference in philosophy. Clearly both methodologies try to establish what outputs from a process are important to quality and then establish which inputs to the process cause a detrimental effect on quality. The difference in language is most likely a result of the Six Sigma development from purely statistical concepts whereas SS development came from an engineering perspective trying to develop more simplified yet equally powerful statistical tools.

This difference in language rather than philosophy extends even further to the frameworks of the two methodologies (see Figure 1 and Figure 2). They both start by defining a project. Establishing the measuring system, although there is a different in achieving this

as SS will use graphical tools such as Isoplots^{SM}, whereas Six Sigma implements statistical measures such as Gage R&R.

The core difference in methodologies seems to be how processes are analysed to identify variation sources. At this stage Six Sigma seems to deviate from its mantra of data driven improvement. In order to identify sources of variation, as previously identified in Section 2.1, subjective methods such as brainstorming and/or cause-and-effect matrices are used to identify the possible sources of variation and statistical tools are used to prove or disprove these hypothesis. Alternatively rather than identifying sources of variation through subjective analysis, complex DoE or correlation plots are used to screen all possible inputs but these methods suffer from confounding, this is something Bhote [11] calls a *"scattergun"* approach. SS differs as a "generate clues" phase is implemented during what will correlate with Six Sigma's analyse phase. During the clue generation SS takes advantage of graphical techniques, such as multi-Vari plots, and statistically valid offline testing tools, such as component search, to eliminate inputs in a process which do not affect the CTQ or Green Y^{SM}. This method of narrowing down objectively to a list of suspect variables allows DoE to be applied with higher-resolution factorials which eliminates or reduces the risk of confounding.

Six Sigma and SS then go through similar actions in the Improve and Control phases which equates to the second half of the SS algorithm, from the Red X being found and optimized.

2.4   Modified Six Sigma Method

This section proposes a Six Sigma/Shainin Hybrid, see Figure 3. Core to the hybrid method is the Six Sigma DMAIC cycle and the Six Sigma 12-step strategy [3]. The Six Sigma methodology was used as the core material as it is overall a strong methodology [7], which has more far reaching and established use in industry compared with Shainin.

SS thinking is used to reinforce the established Six Sigma methodology in two ways, firstly to include a "find the signature" and "narrow down" loop to the identification of variation sources. The philosophy of using graphical techniques to eliminate factors which definitely don't affect CTQ's increases the objectivity and expands the "exploration" of a problem process during the analysis. The second SS loop is the addition of the "switch on/off" and "optimise" to the improve phase. This addition aims to utilise Shainin tools such as Variables Search^{SM} to screen the remaining possible factors in a concise manner without missing interaction effects. These specific techniques will be described and

discussed further in later chapters. This method should still capture all the normal Six Sigma Metrics such as sigma level/Process Capability, Gage R&R, etc.

```
                                    ┌─────────┐
                                    │ DEFINE  │
                                    └────┬────┘
─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─│─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─
┌──────────────────┐                         ┌──────────────────────┐
│                  │                         │ Select Critical To   │
│ • Performance    │                         │ Quality              │
│   Metrics        │              ┌─────────┐│ Characteristics      │
│ • Gage           │              │ MEASURE ││                      │
│   Validation     │              └─────────┘│ Define Performance   │
│   Studies        │                         │ Objectives           │
│ • Isoplot        │                         │                      │
│                  │                         │ Validate Measurement │
└──────────────────┘                         │ System               │
                                             └──────────────────────┘
```
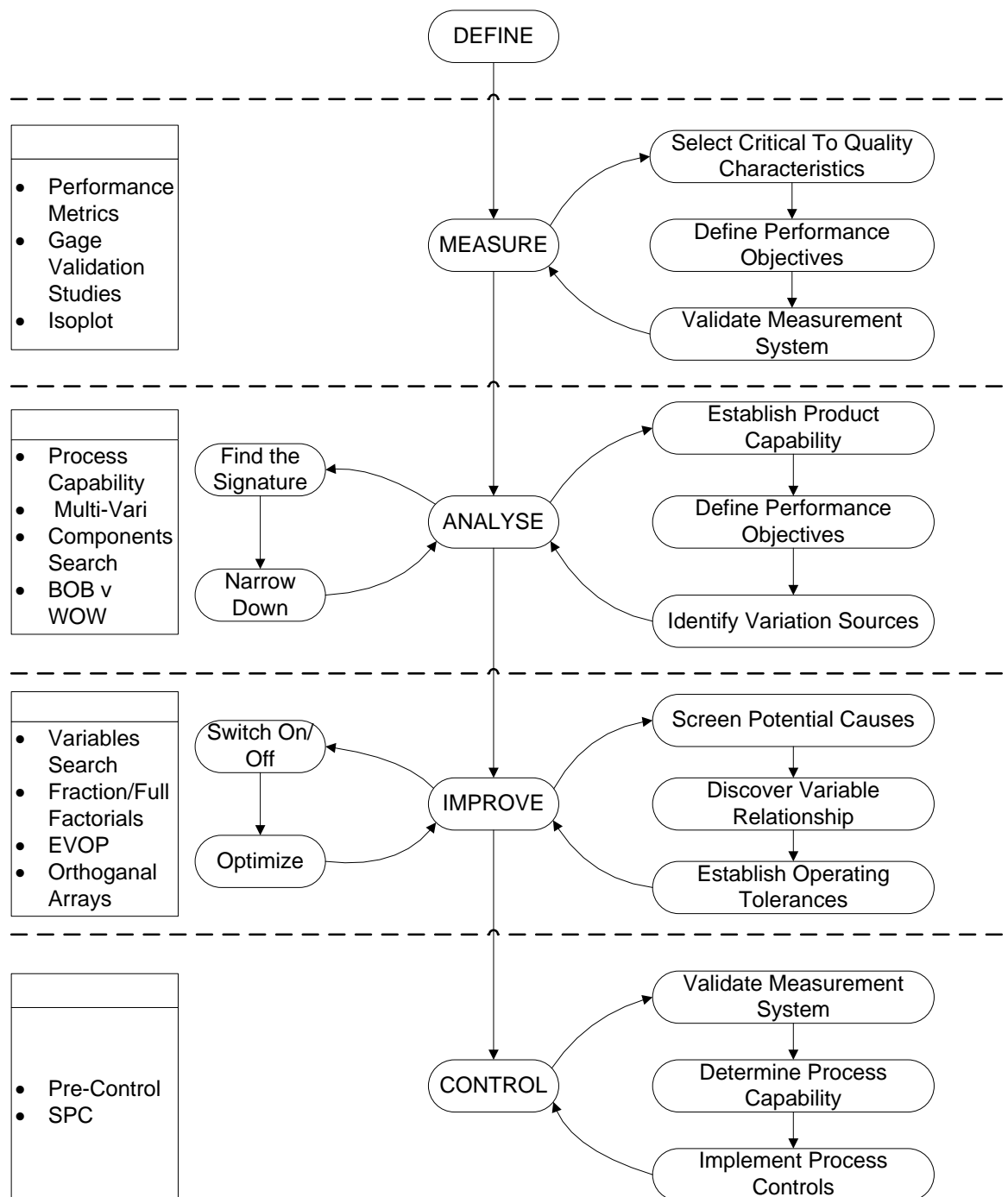

Figure 3 Hybrid Six Sigma Shainin Framework

The principle of Hybrid Six Sigma methodologies is not new; George [6] offers one view on Lean Six Sigma. This introduces Lean manufacture principles to the DMAIC process

8

improvement cycle. Given the historical development of Six Sigma and SS it is less of a leap to see a hybrid of them. Shainin problem solving tools and SS have been developed and adapted by Dorian Shainin and later his consultancy firm since 1943 [8] independent of Motorola were Six Sigma was conceived. However Shainin tools must have been used within Motorola, Shainin's main advocate (and *"self-confessed disciple"* of Shainin) Keki Bhote had a 42 year career with Motorola and contributes the publication of Shainin techniques in [11] to Motorola winning the Malcolm Baldrige National Quality Award in 1988. Yet there is no mention in Six Sigma texts about using Shainin tools, this is probably a result of the proprietary nature of Shainin tools [2] and the *"hyperbole"* which Bhote uses putting many off from using them [10].

The described Six Sigma Shainin methodology will now form the framework that further process improvement techniques and teaching materials have been developed to operate within. The next chapter utilizes this framework to develop a concise process improvement tool that bridges the measure/analyse phase validating the measurement system and establishing capability, whilst taking advantage of Shainin Techniques to find the signature of the variation.

3   Process Variation Diagnostic Tool

This chapter will outline a sampling strategy known as the Process Variation Diagnostic Tool (PVDT). The PVDT was devised out of the need to improve the objectivity of the early analysis of a quality problem when there are a large number of factors in a process to analyse and a relatively low volume of product to sample, either due to time, cost or the volume of production. Reducing the time spent on a quality problem was a major consideration in the design of the PVDT. It was realised that prerequisites to fulfil the DMAIC cycle are a Gage R&R study and a Process Capability study, techniques classically used in the Six Sigma Measure and Analyse phase respectively. These techniques take time to apply and unless there is a measurement system problem will shine little light on the root cause of a problem.

The PVDT will provide a method whereby from 20 samples a Gage R&R and a Provisional Process Capability study can be found. It is structured so it will also provide an Isoplot^{SM}, which will give a graphical representation of the reproducibility of the measurement system and a Shainin Multi-Vari study a technique associated with the "*Clue Generation*" phase of SS [9]; from which the "*signature of variation*" can be found using data collected to validate the measurement system in the Measure phase of the DMAIC cycle. This allows for a more efficient analysis of a process problem, as it starts the Analyse phase by reducing down the numbers of factors under consideration by eliminating unimportant factors objectively with data driven information. This reduction in factors by a Shainin Multi-Vari analysis significantly reduces the subjectivity of the early analysis when compared to previously discussed techniques such as Cause and Effect Matrix or Brainstorming. Thus later analysis with DOE can be more powerfully and more meaningfully applied to find the underlying effects of the important factors. As fewer important factors to analyse with DOE, the fewer experiments are needed to fully understand the interaction effects between them. The Shainin philosophy is best surmised by Bhote [11] stating *"don't let the engineers do the guessing, let the parts do the talking"*.

3.1   Gage R&R: A Test of the Measurement System

A Gage R&R study is a generic Six Sigma term for measurement system analysis [15]. In the manufacturing area the Gage R&R is the most common test of the effectiveness of a measure [4].   It involves repeating a measure with different appraisers or measuring equipment to test against the repeatability and reproducibility of a gage. The Gage R&R in the PVDT will follow criteria set out in the QS9000 reference manual, Measurement System Analysis [16].

Part of the PVDT approach is to outline a "roadmap" of follow up techniques that fit with the Hybrid Six Sigma methodology previously outlined (these are discussed later) to find the important factors which contribute to the root cause of a quality problem. This is important as Dasgupta [17] found on many occasions that Gage R&R studies are being restricted to a simple evaluation of the measurement system with the object of satisfying a third-party auditor and no action following. It is therefore important that when a PVDT is undertaken it is done with suitable resources available to perform follow up studies of any measurement systems deemed inadequate.

3.2   Provisional Process Capability Study: Analysing Process Performance

A Process Capability study aims to quantify statistically the variation with which a process produces products compared with the specified tolerances. It is common in the Analyse phase of Six Sigma, to establish the current process capability which acts as a baseline for improvement projects (a complete description is in the QS9000 reference manual, Statistical Process Control [18]). At this stage the results should not be used to protect future performance as it is yet to be determined if the process is under statistical control. When data is collected and analysed, using statistical software such as Minitab or plotted on normalised graph paper, this information can also be used to demonstrate normality graphically.

A further discussion is whether the capability calculated by the PVDT is a short term index denoted by C or a long term index denoted by P. In terms of the calculation of C_p or C_{pk} against P_p or P_{pk} the formulas used are the same, the issue is a technical point over whether the data collected can be considered long term. The first issue is finding a definitive definition of long term in six sigma texts as most brush over the issue. However George [6] states *"'long term' means data have been collected over a long enough period that you believe it likely you have seen 80% of the process variation."* If this is true the PVDT should apply the P_p and P_{pk} indices as a pre-requisite of the Shainin Multi-Vari is that the samples are collected over a long enough time frame to capture 80% of the process variation. Therefore the overall capability measured by the PVDT is denoted by the P metric.

The use of the P_p study is limited to cases where there is a double sided specification present. This is a metric which compares the process variation with the specified process tolerance, it is therefore a useful measure of a processes performance but it does not consider whether the process is centred. P_{pk} studies however can be applied to both single and double sided specification and are influenced by process variation and centring. It is therefore recommended that in the case of a double sided specification both P_p and P_{pk} are

measured to determine whether poor process performance is a result of process variation or process centring. In the single sided case using only P_{pk} will indicate the defect rate of the process regardless of the width of process variation.

3.3 Isoplots[SM]: Analysing an Inadequate Measurement System

This is a graphical technique which can be used to home in on whether a poor Gage R&R is a result of product variation, poor repeatability, poor reproducibility or whether the test process itself is having an effect on the product.

Isoplots[SM] are constructed by taking a sample of 30 units from a process. Testing 15 units on test equipment 1 and then retesting them on test equipment 2. The second batch of 15 units are tested on test equipment 2 and then again on test equipment 1. Plotting the results on a graph where both axes have the same scale and for each unit the results of test equipment 1 on one axis and the results of test equipment 2 on the other axis, should show one of three things; if the results are spread along a 45° line, this shows product variation, as shown in the example in Figure 4. If the results are spread in a direction not on a 45° line this shows poor measurement variation or poor reproducibility. If the product variation is five times greater than measurement variation this will give the test 98% confidence in its measurement system. If the results are in two groups either side of the 45° line this shows that the test process has an effect on the unit. Isoplots[SM] can be performed on the same test equipment using different operators to show if there is a reproducibility problem between operators.
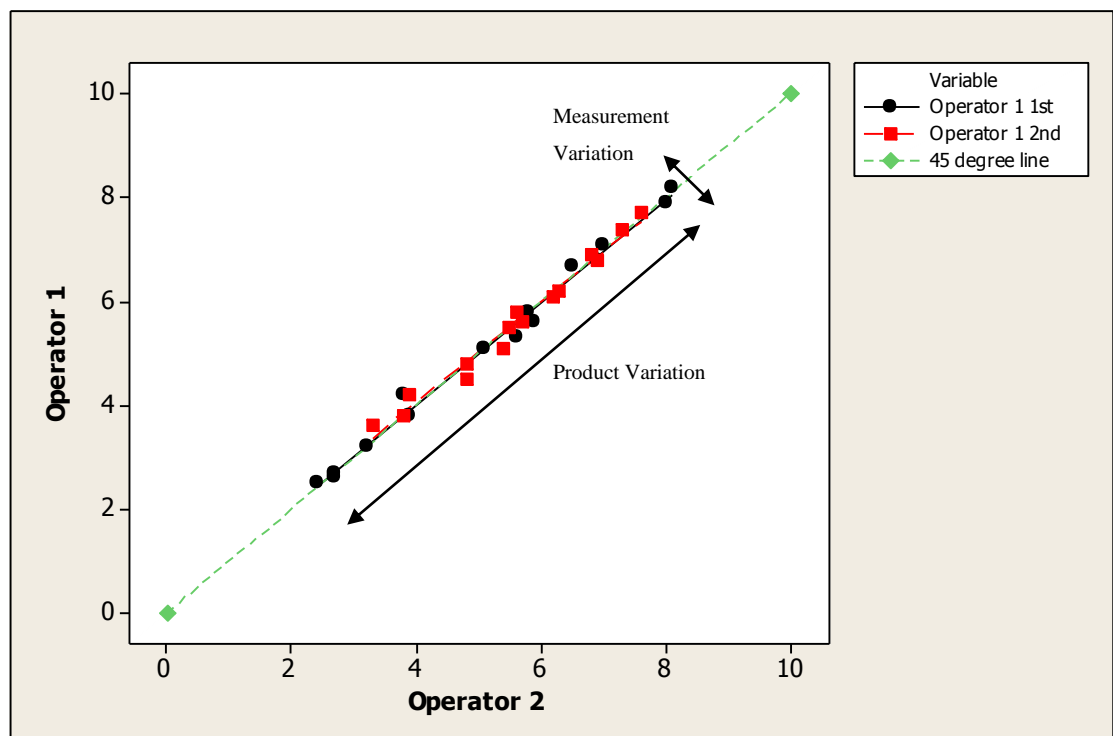
Figure 4 Example of a Isoplot, showing the difference between measurement and product variation

12

3.4 Shainin Multi-Vari: Finding the Signature of Variation

The use of Multi-Vari studies as a method of separating types of variation is described in Seder [19] were the credit for the technique is given to Dr. Joseph M. Juram. Although Dorian Shainin did not originally conceive the method and acknowledges this in his forward to Bhote [11], in this section it is referred to as a "Shainin Multi-Vari" as the sampling constraints devised by Shainin are used in this thesis. The use of the term "Shainin Multi-Vari" is also seen in the Help file of Minitab version 13 stating "*Minitab draws Shainin multi-vari charts.......a way of presenting analysis of variance data in a graphical form providing a "visual" alternative to analysis of variance.*"

A Shainin Multi-Vari study is described in Bhote [11][12] Steiner [10] and De Mast [20] is used to find the "signature of variation", categorized in "Shainin speak" as the Red X^{SM}, Pink X^{SM} and Pale Pink X^{SM}. These causes are visually displayed in a Shainin Multi-Vari study as one of three special families; within-piece, piece-to-piece and time-to-time. These families have specific types of causes associated with them and the family which is categorized as the Red X^{SM} should be investigated first. This philosophy is based on the Pareto principle that the vital few causes account for the majority of a quality problem, therefore eliminating your Red X^{SM} should have the most effect on your variation.

Within-piece variation occurs within a single unit due to a poor measurement system or is a result of non-uniform product. Piece-to-piece variation occurs between consecutive units or within groups of four units, due to individual processes, random variation or within-piece variation at a different level. Time-to-time variation occurs between groups of four products, due to hour-to-hour, shift-to-shift or batch-to-batch changes. Figure 5 is an example of a Shainin Multi-Vari, study showing how the study is used to graphically display the families of variation.

In its basic form the Shainin Multi-Vari Study provides a visual display of the size of the signature of variation. This field of inquiry can be reduced and a suitable experiment can be performed depending on the signature of variation. This is a technique which does not rely on complex statistics and it can quickly narrow down the search for the Red X^{SM} without resorting to guesswork. The visual display can if necessary be supplemented by an analysis of variance (ANOVA) to numerically estimate the size of each family of variation as proposed by De Mast [20]. This extra analysis does add statistical confidence to the categorization of the Red X^{SM} but also adds complexity and time which is only necessary when two families show similar amounts of variation.
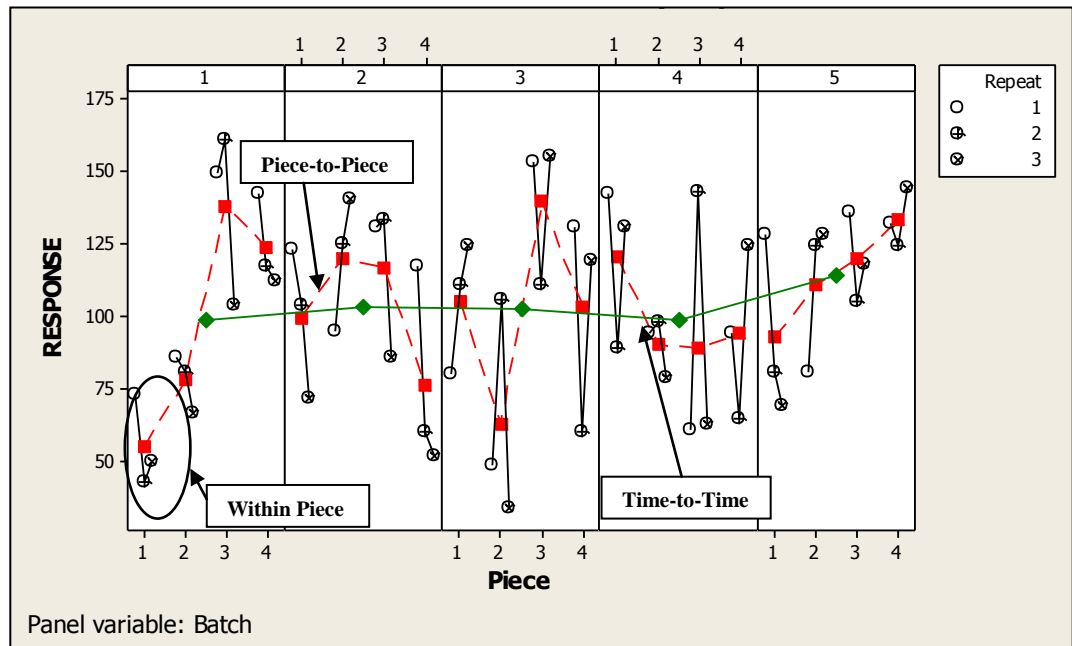
Figure 5 Example of a Shainin Multi-Vari, demonstrating the three Signatures of Variation

This section will now be divided into two further parts firstly describing the implementation of the PVDT and a description of three case studies which have benefited from the use of it.

3.5 Process Variation Diagnostic Tool Methodology

The PVDT is an example of how Six Sigma tools, at the interface of the Measure and Analyse phases of the DMAIC quality improvement cycle, and Shainin tools, at the clue generation phase of the Shainin System algorithm, can be used in the Six Sigma DMAIC framework simultaneously to achieve effective data driven improvement. The approach differs from pure Six Sigma and Shainin in the following ways:

1. Although it is structured in a way to provide a Shainin Multi-Vari, it is also structured to provide a Gage R&R and Provisional Process Capability Study. These are essential for the Measure and early Analyze phases of Six Sigma.

2. Gage R&R and Provisional Process Capability studies also add numerical information to the graphical Multi-Vari analysis. Gage problems or Within-piece problems seen in the Multi-Vari study are quantified as is the overall variation seen in the Multi-Vari study. Also the Isoplot[SM] test will highlight if the test is effecting the measurement very clearly, something the numerical data of the Gage R&R will not show.

3. Rather than seeing Six Sigma and Shainin as competing methodologies, it has introduced the philosophy of narrowing down to important factors (rather than indentifying them early), into the Analyse phase of the DMAIC cycle.

14

The advantage of the PVDT is that from a sample of 20 products four statistical techniques can be conducted from the same results. Cutting down the time spent on collecting information for these procedures. Also the use of an Isoplot^SM and a Shainin Multi-Vari study will add significantly to the early analysis at no extra cost in terms of time or samples needed when compared to classic Six Sigma.

3.6 Implementation and Sampling Structure

- From a previously defined problem process collect 20 sample products.
- The samples should be collected in groups of four either consecutively from a flow line or, when products are produced in batches, from across the batch. This should be done at five separate times providing 20 samples.
- The five measurement times should be selected so that 80% of the variation normally found in a process is captured; this can normally be predicted from historical information used to Define the project. This could be over a shift, a day, a week or if there is batch production, the 5 periods could correspond to five batches.
- The critical-to-quality characteristics (CTQ) on the samples should be measured three times. The first 10 sample measured twice by appraiser/measuring equipment 1, then once by appraiser/measuring equipment 2 and the second 10 samples measured twice by appraiser/measuring equipment 2, then once by appraiser/measuring equipment 1
- The results are then analysed using the following techniques; Gage R&R, Isoplot^SM, Process Capability and Shainin Multi-Vari.

3.6.1 *Unbalanced Gage R&R by the Xbar and R Method*

Measuring 20 samples three times, twice by one appraiser or equipment and once by a second appraiser or equipment, we are able to conduct a Gage R&R study by adapting the average and range method, set out in [16], to have an unbalanced design as follows:

1. Find the average range between operator one first and second measurements $(\overline{R_e})$
2. Find the average result for operator one's second measurements $(\overline{X_1})$
3. Find the average result for operator two's measurements $(\overline{X_2})$
4. Calculate the repeatability or equipment variation (EV):

$$EV = \overline{R_e} \times \frac{5.15}{1.128}$$

5. Calculate the reproducibility or appraiser variation (AV):

$$AV = \sqrt{\left((\overline{X_1} - \overline{X_2}) \times \frac{5.15}{1.41} \right)^2 - \left(\frac{EV^2}{40} \right)}$$

6. Calculate the repeatability and reproducibility(R&R):

$$R\&R = \sqrt{EV^2 + AV^2}$$

7. Finally calculating the R&R value as a ratio against tolerance:

$$\frac{R\&R}{USL - LSL} \times 100\%$$

Where USL is the upper specification limits and LSL is the lower specification limits.

With the QS 9000 guideline suggesting a value of less than 10% as adequate, between 10 and 30% as marginal and anything greater than 30% as inadequate.

3.6.2 Applying Process Capability to Establish a Performance Measure

Using the first measurements taken by appraiser/measuring equipment 1 in the Gage R&R we are able to conduct a provisional process capability study as defined by QS9000:SPC [18]. It is called "*provisional*" for two reasons, it is taken near the beginning of the process improvement cycle and is therefore a benchmark for gains found during the improvement initiative. It is also obtained using 20 samples rather than the normal 50 samples; therefore this Process Capability Study should be used as benchmark for a project rather than a method of protecting future performance as it is not being performed in conjunction with control charting techniques for example to establish statistical control. Deleryd [21] gives a useful insight into the practical use of process capability based on surveys of companies using process capability metrics, acknowledging that these studies can improve process knowledge and identify improvement opportunities based on data. However disadvantages identified are resource consuming and there is a potential for error if they are misused or given too much credence if statistical control of a process is not established.

Particular care must be taken when deciding the correct approach for this calculation, as it will depend on whether the results are attribute or variable measures and if the results have a double-sided or single-sided specification. The most desirable situation which is common in occurrence is a variable measure which is distributed normally. In this situation it is typical to calculate P_p for a double-sided specification and P_{pk} for both a double or single-sided specification as follows:

$$P_P = \frac{USL - LSL}{6\sigma}$$

$$P_{PU} = \frac{USL - \bar{X}}{3\sigma}$$

$$P_{PL} = \frac{\bar{X} - LSL}{3\sigma}$$

$$P_{pk} = min[P_{pu}, P_{pl}]$$

Where:

LSL is the lower specification limit

USL is the upper specification limit

σ is the standard deviation

$\bar{X}$ is the sample mean

When there is a double-sided specification the P_p capability index is a measure of the width of process variation compared with the allowable variation or tolerance. P_{pk} capability index also factors in how well centred the process is against the process target.

3.6.3 Plotting an Isoplot^SM to Graphically Display Measurement System Variation

The first results for operator or measurement equipment one should be plotted on the same scale against the results for operator or measurement equipment two, as in Figure 4. The graph can then be assessed as previously described in section 3.3.

3.6.4 Plotting a Shainin Multi-Vari using the Data Collected

The Multi-Vari chart should be plotted as in Figure 5, whereby the y-axis is labelled with the measurement values and x-axis is labelled with the product numbers. Each product should have three values with lines drawn between the maximum and minimum values to graphically represent the within-piece variation. The mean measurement value for each product should then be joined up in groups of four to represent the piece-to-piece variation. The five groups of four products means should then be joined to show the time-to-time variation.

3.6.5 Using Minitab to Automate the Analysis

As part of the research a macro was developed to automate the analysis of the 60 results from the 20 samples using Minitab. This Macro can be seen in Appendix A: Minitab Macro. Minitab software was used due to its wide use in organizations which employ Six Sigma, it was therefore important to make the method compatible with Minitab.

The macro utilizes the built in features of Minitab to conduct a Process Capability, giving both within or C indices and overall or P indices, and construct a Shainin Multi-Vari Study. The difficulty in the Macro was enabling Minitab to conduct an Unbalanced Gage R&R study as previously outlined. Minitab does not process Unbalanced Gage R&R designs using the built in Gage R&R package. The macro works round this by using the built in Gage R&R feature twice, using the Xbar and R method, firstly to calculate the unadjusted reproducibility (UAV) and then to calculate the repeatability (EV). The macro then finishes the calculation using the QS9000 guidelines to adjust the reproducibility value and then find the R&R value as a ratio against the tolerance.

3.7 Analyzing and Following up the Results

If the Gage R&R study shows a measurement system to be inadequate the focus of the follow-up investigation should be to find the cause of this variation. This type of variation will also show on the Multi-Vari study as within-piece variation, it may not however be the Red X^{SM}. The Isoplot^{SM} should demonstrate whether this variation is a result of faulty test equipment, operator error or if the test process itself is having an effect on the product. This variation could also be a result of product variation (which would be established by the Isoplot^{SM}) for example a shaft which is out-of-round, in which case a Concentration Diagram could be used to home in on the " root cause" of variation.

If the Gage R&R study shows no measurement system problem or any measurement system problem has been resolved, the Shainin Multi-Vari study is used to determine the next biggest family of variation; Figure 6 summarizes the follow-up investigations which can be applied once the signature of variation is found.
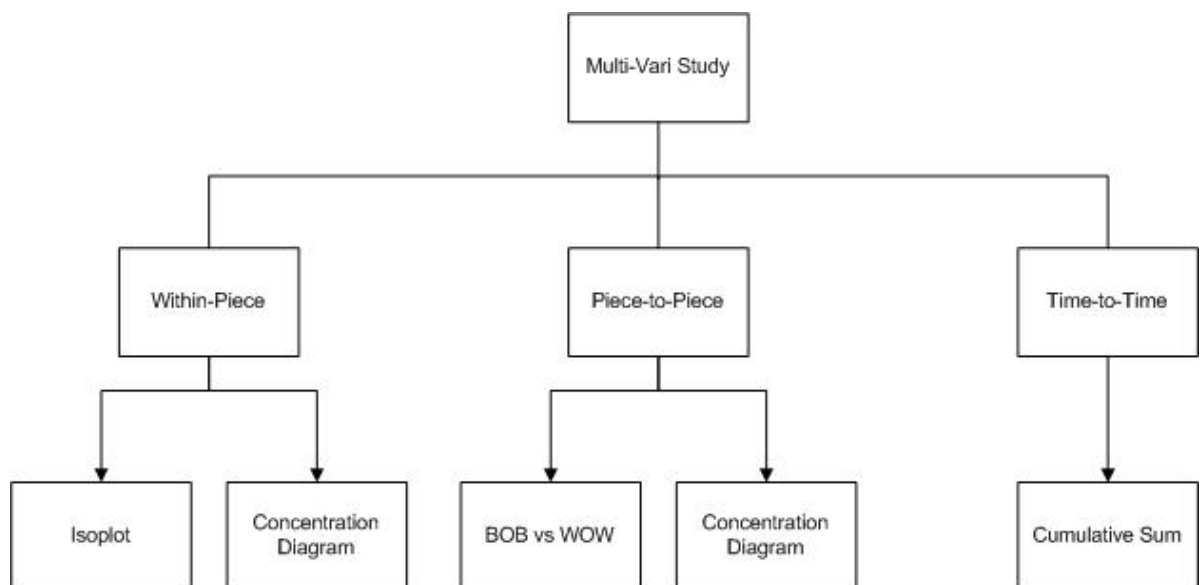
Figure 6 PVDT Roadmap of techniques to guide the follow-up investigations

Piece-To-Piece variation can be followed up with a Best of the Best vs Worst of the Worst (BOB vs WOW) investigation to observe the differences in a process between good products and bad products. A Concentration Diagram could be used to observe if there is a within piece problem higher up the process.

Time-to-time variation can be followed up with a cumulative sum technique to observe the exact point when there is a change in a product and to link this change in product to a change in process. After these investigation DOE such as full factorials can be applied more efficiently as the number of factors involved in the DOE has been significantly reduced by the previous investigations.

The following section will now give case study evidence the PVDT offers an effective tool for driving forward process improvement projects in a concise manner. Section 5 will then detail the most recent case were the PVDT was used by the author in a live industrial case.

4   Case Studies of the Practical Implementation of the PVDT

This section will outline three case studies where the PVDT has been implemented in industry. The first two studies were carried out by 2 people over a period of 2 weeks. The third study was conducted at a company with a chronic quality problem which had been unresolved after 12 months of investigation.

4.1   Case One: Furniture Manufacturer

The first case was conducted at a leading furniture manufacturer. The site specialised in taking in chipboard and MDF as raw materials, then cutting them to size and finishing with wood effect foils to form finished panels ready for assembly. Figure 7 shows an overview of this Edge Banding process.

The process of cutting and applying glued strips of wood effect edging to the chipboard panels is a largely automated process. With all four edges being processed in one run, Figure 7 shows how the long edges processed first. They are sawn to the required width; the edge banding is applied and then trimmed level with the panel. The panel is then rotated and the short edges then go through the same process, except they go through an additional snipping process to ensure the edging applied is the correct length, before the panel is inspected. Output from these machines is around 10,000 panels per day. The outputs of these machines had been subject to a long standing quality problem and had been subject of a number of process improvement programmes over the years. At the start of the study the company was seeing around 20% of its output being returned due to edging problems.
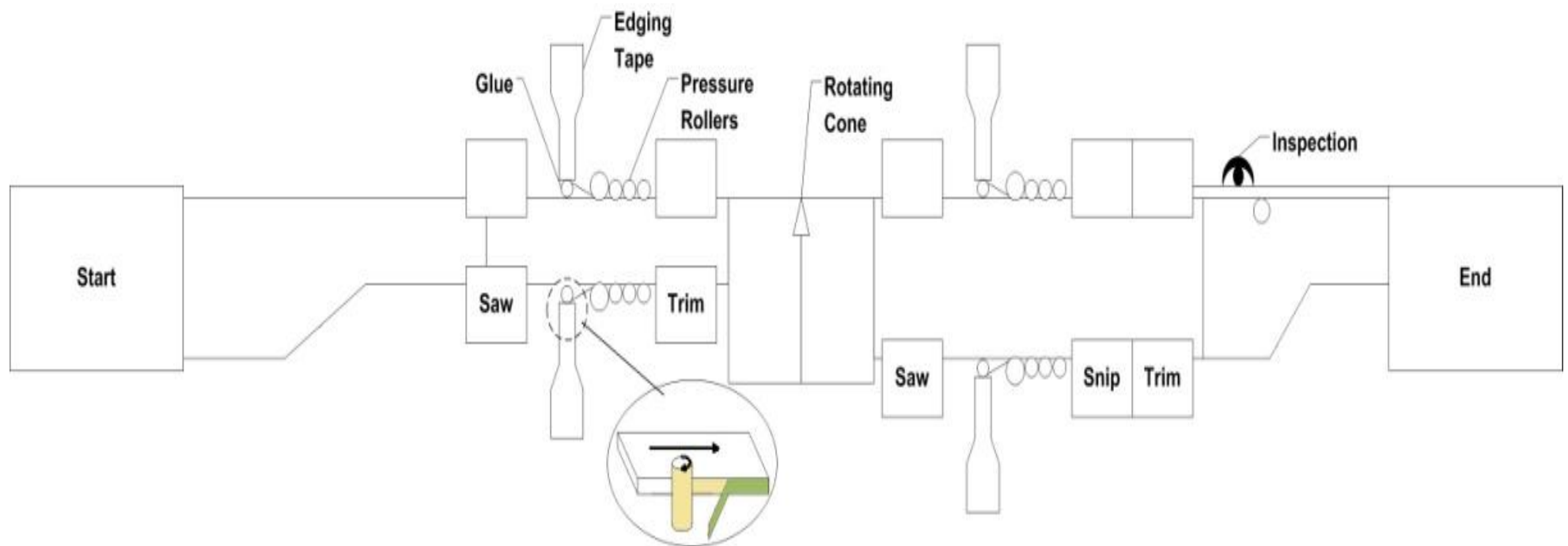
Figure 7 Overview of the edge banding process showing the layout of the tooling

4.1.1 PVDT Implementation

The project team were able to quickly define over and under trimming of the edging as the focus of the process improvement programme. In order to determine the signature of this variation a PVDT was set up with the following constraints; four consecutive panels were acquired from five different time periods, each edge was measured three times. This means from a sample of 20 panels, each edge has a total of 60 measurements taken and a total of 240 measurements taken around all four edges. Each of the four edges had the PVDT applied separately as they were trimmed at different points in the machine process. Thus each edge was numbered, as in Figure 8, to catalogue the results separately.

1 Long Edge Top Side

4 Short Edge Bottom Side

3 Short Edge Top Side

2 Long Edge Bottom Side
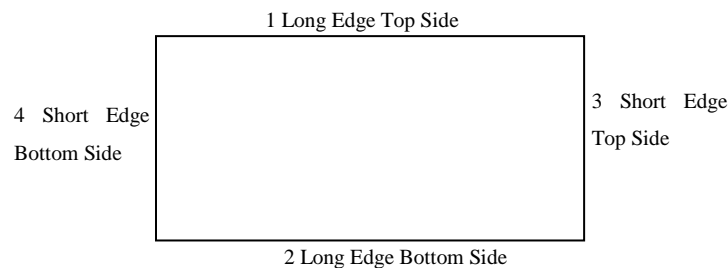
Figure 8 Panel Edge Labelling to identify the machine responsible for each edge

Finally consideration had to be made as how to implement a quantitative grading score for the problem, as prior to this project the product was considered as acceptable or rejected. The following grading system (Table 1) was implemented;

Table 1 Modified Labeling System Implemented

2	Reject	(very under trimmed, out of specification)
1	Acceptable	(under trimmed, within specification)
0	Good	
-1	Acceptable	(over trimmed, within specification)
-2	Reject	(very over trimmed, out of specification)

4.1.2 Gage R&R Results to test the measurement system

The results shown in Table 2 demonstrate that there was serious problem with either the measurement system or non-uniformity along the edging. Three of the results are above 30% with classes them as inadequate and one result is greater than 10% which classes it as marginal.

Table 2 Gage R&R Results from the Edge Banding Process

Panel Edge	% Gage R&R
1	78
2	80
3	53
4	20

4.1.3 Provisional Process Capability study to determine the process performance

The capability studies in Figure 9, Figure 10, Figure 11 and Figure 12 show a $C_p \leq 1$ in all cases, which is extremely low indicating the process, is failing to produce sufficient products within specification. Which is consistent with the high numbers of product returns that was experienced prior to the investigation.
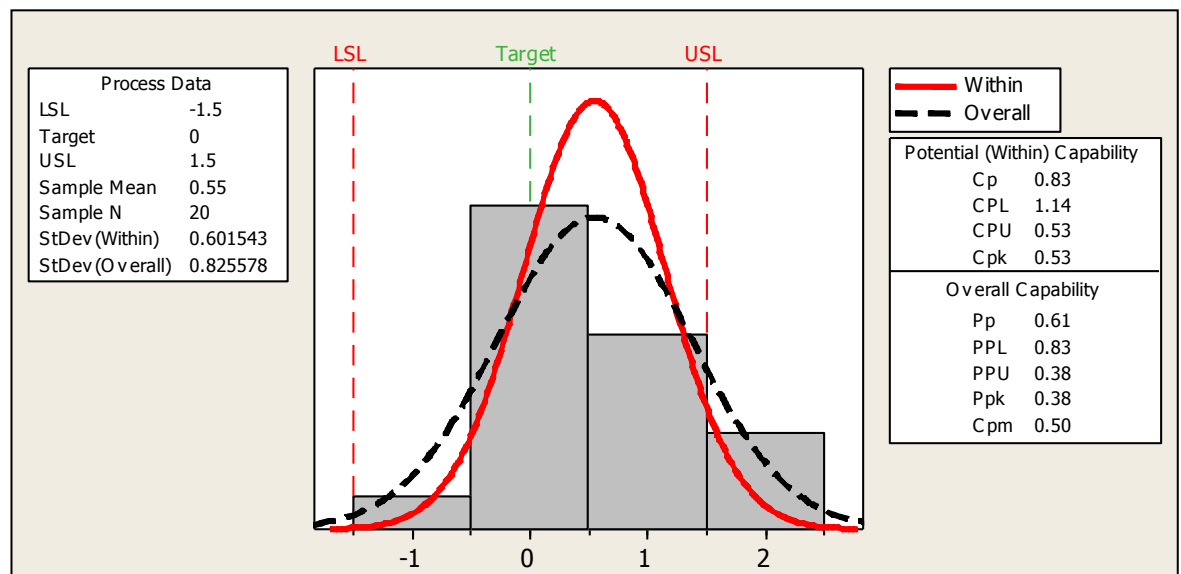
Figure 9 Provisional Process Capability for Edge 1, with Cp < 1 showing the process is not very capable
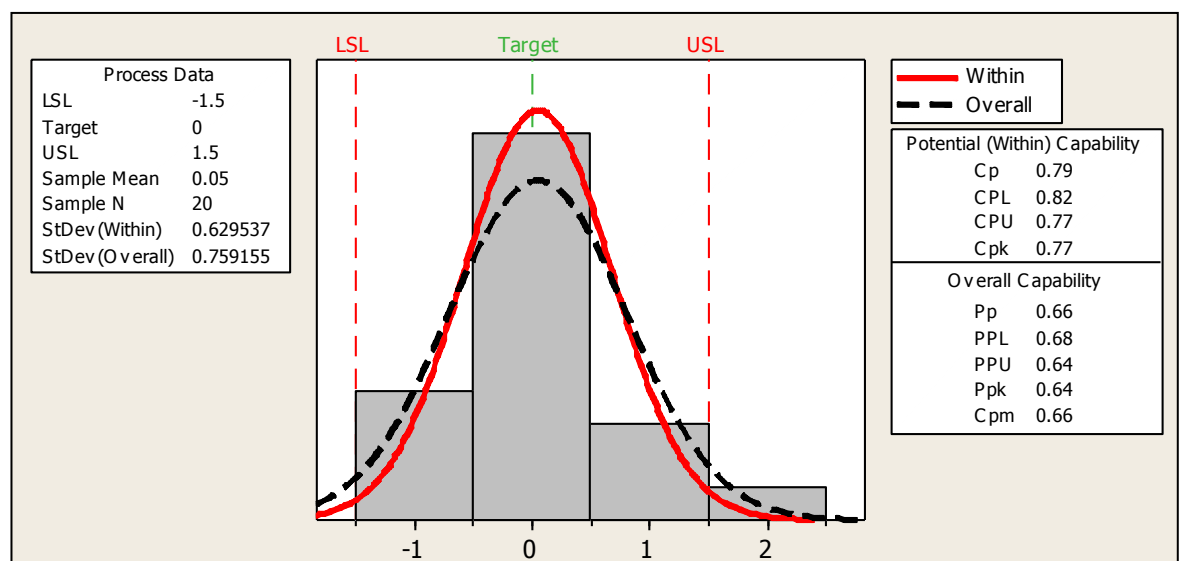
Figure 10 Provisional Process Capability for Edge 2, with Cp < 1 showing the process is not very capable

23
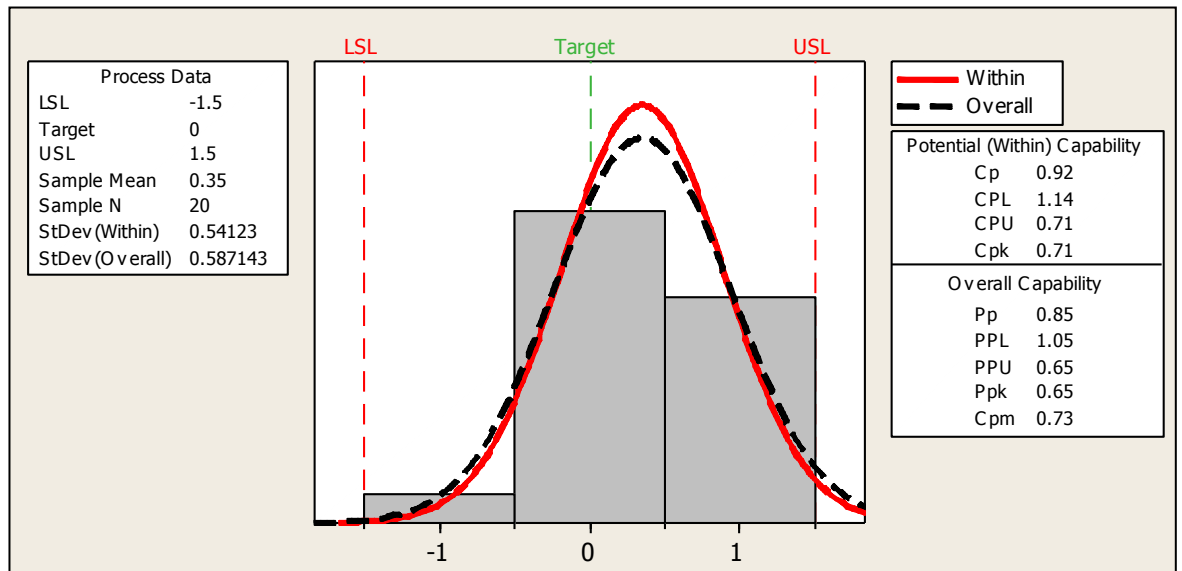
Figure 11 Provisional Process Capability for Edge 3, with Cp < 1 showing the process is not very capable
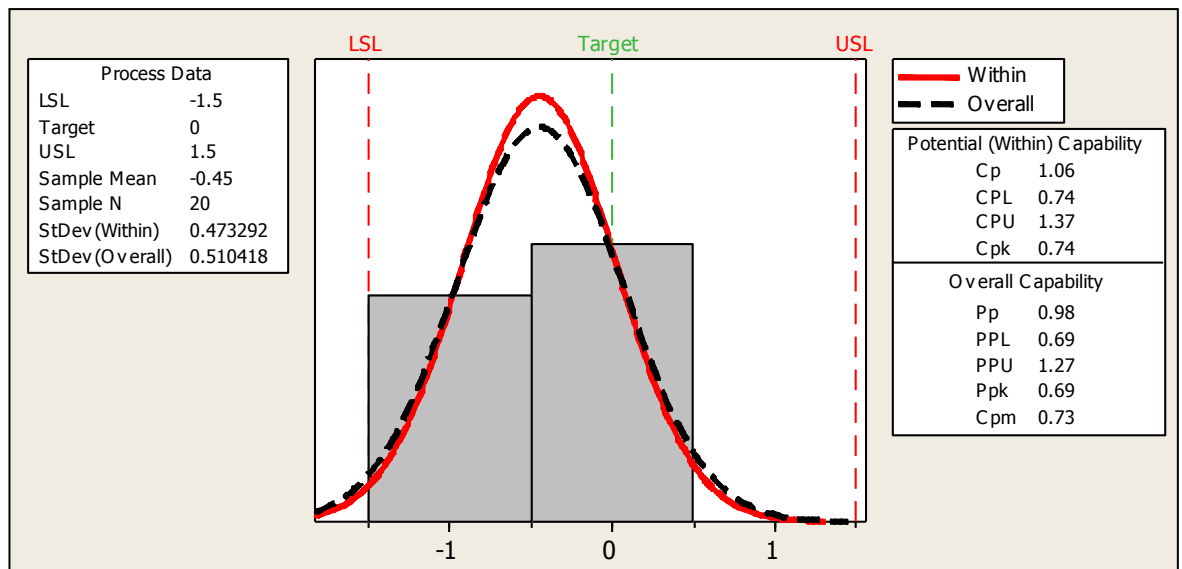
Figure 12 Provisional Process Capability for Edge 4, with Cp ≈ 1 showing the process is not very capable

4.1.4 Shainin Multi-Vari Study finding the signature of variation
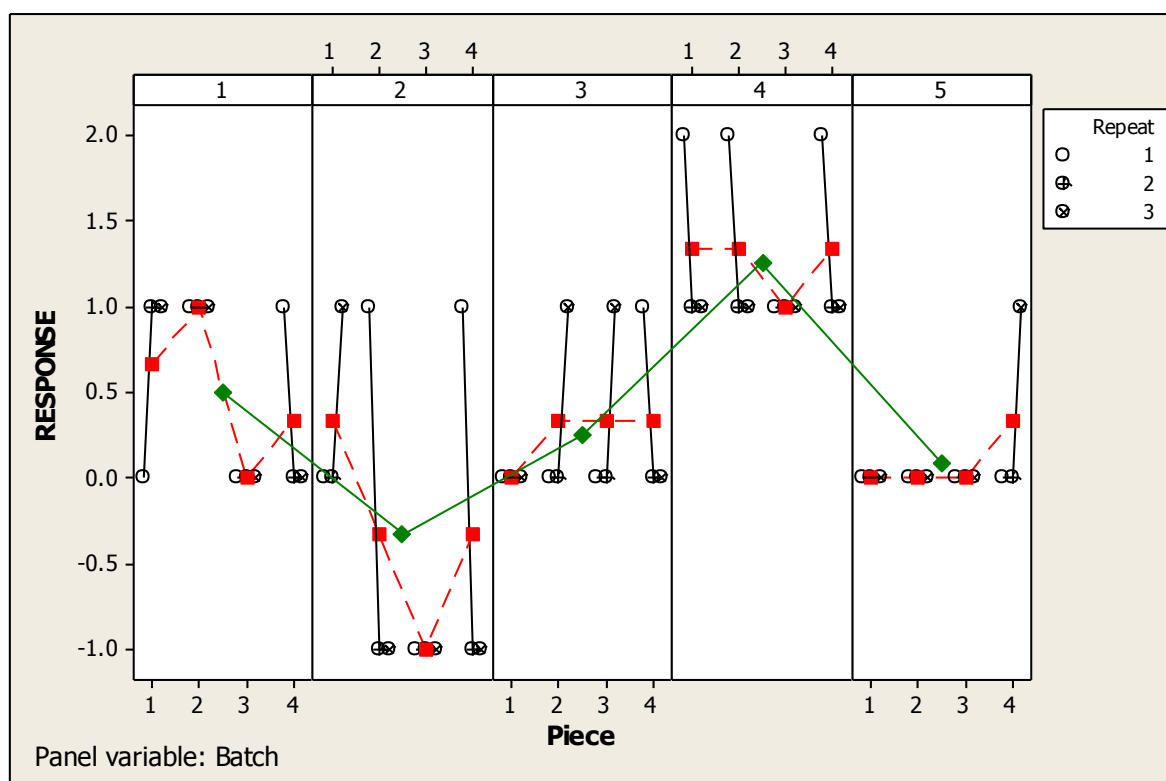
Figure 13 Shainin Multi-Vari Study for Edge 1, showing Red X^{SM} as within-piece and a possible Pink X^{SM} Time-to-Time
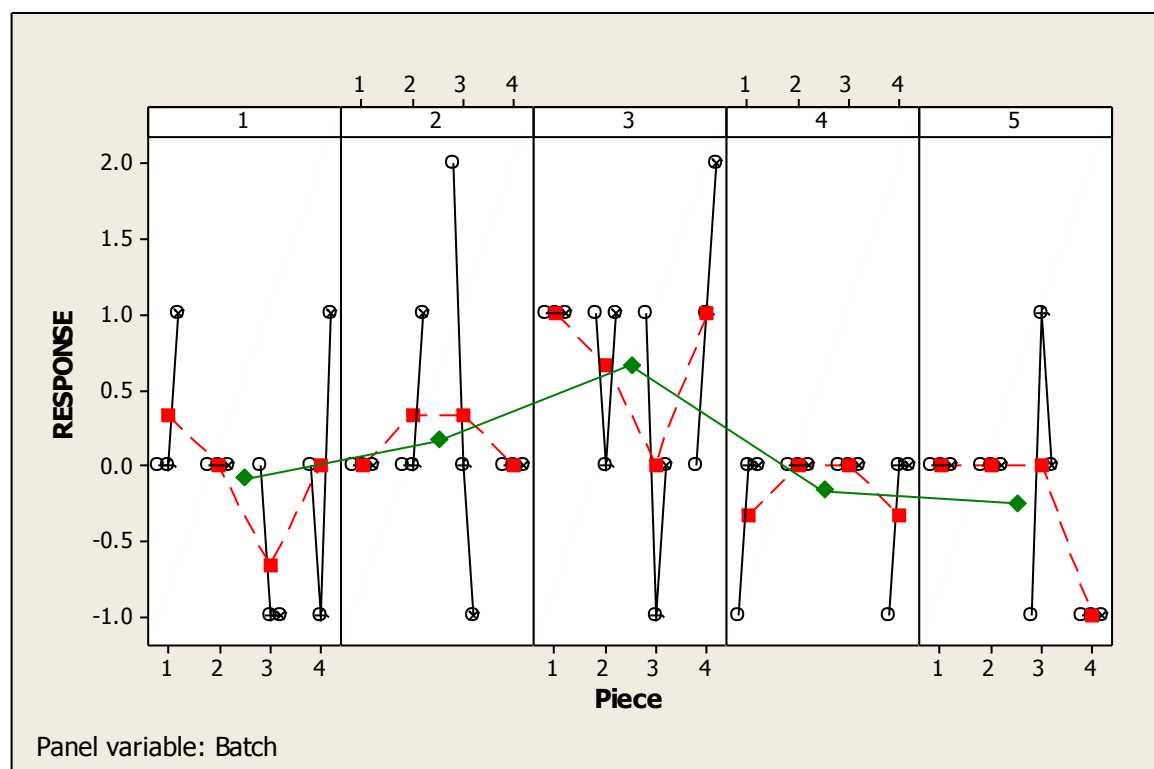
Figure 14 Shainin Multi-Vari Study for Edge 2, showing Red X^{SM} as within-piece

The Shainin Multi-Vari investigations, in Figure 13, Figure 14, Figure 15 and Figure 16, clearly showed the Red X^{SM} signature of variation was predominantly a within piece problem. This was consistent with the Gage R&R study which highlighted large variation

across measures of the same piece. Although the Shainin Multi-Vari shows there is variation piece-to-piece and time-to-time, it also demonstrates finding the factors responsible for the significant within-piece variation will have the biggest effect on improving the capability of this process.
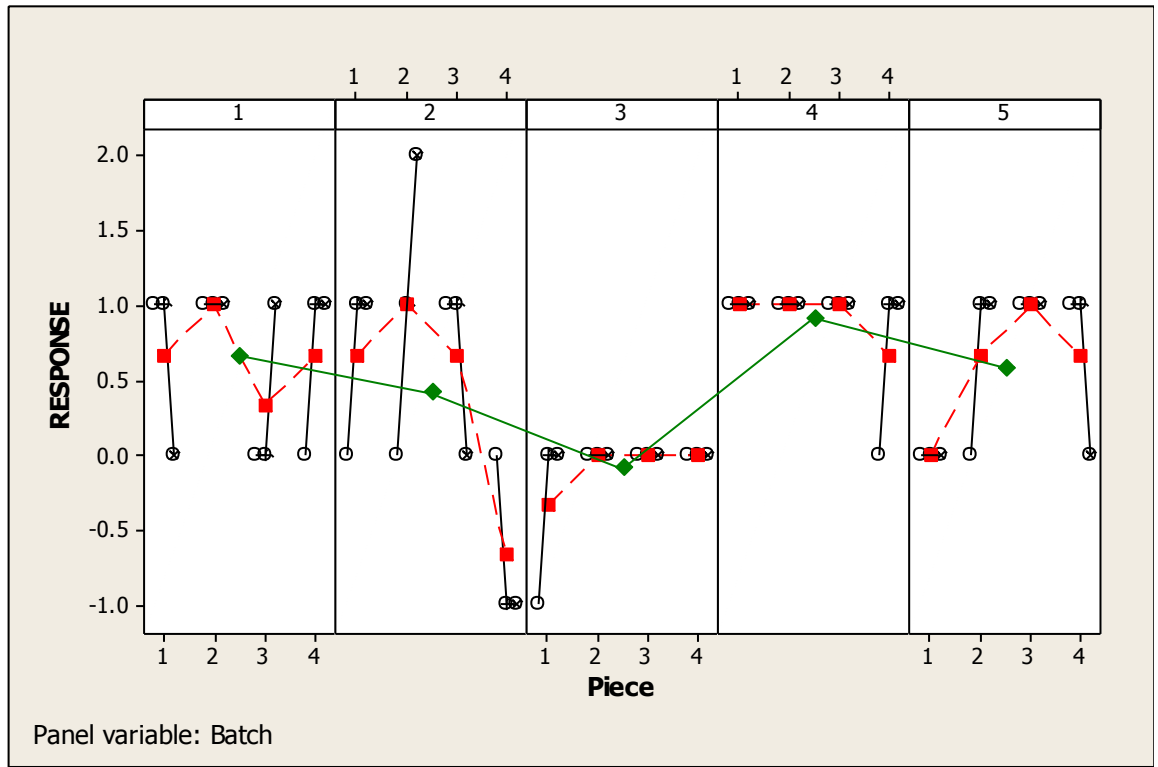
Figure 15 Shainin Multi-Vari Study for Edge 3, showing Red X^{SM} as within-piece
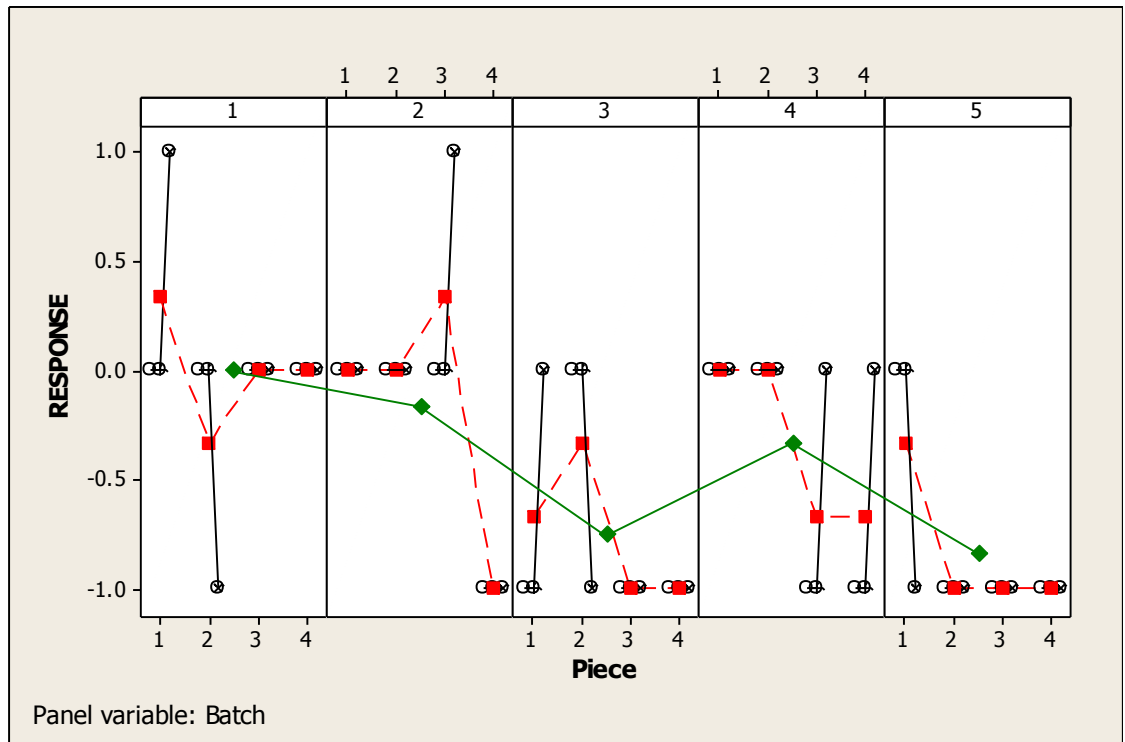
Figure 16 Shainin Multi-Vari Study for Edge 4, showing Red X^{SM} as within-piece and a possible Pink X^{SM} Time-to-Time

4.1.5 Conclusions of the Edging Case Study

From the PVDT the following previously suspected factors were ruled out of the investigation:

- Different size panels are affecting trimming performance; if the edging and trimming machines were affected by the size of the panels there would be a significant batch to batch change in variation. As the sizes of panels being trimmed were being changed between batches.
- Settings are being altered between batches; again the effect of changing setting to accommodate different size panels would show up as a batch-to-batch problem

From the PVDT the future investigation could be focused on;

- Full validation of the measuring system using Isoplots^{SM}; to ensure large variation due to a poor measurement system isn't masking another problem.
- If the measurement system is found to be accurate, the investigation should focus on the edging and trimming machines. Worn parts in a trimmer, for example, could account for uneven edging that result from over and under trimming.

Unfortunately the follow-up investigations were concluded in house outside of the scope of the 2 week project, but the in-house Quality Engineer commented that the project team had *"driven the project further in 2 weeks than it had been in the previous 2 years"*.

4.2 Case Two: Electronic Control Systems Manufacturer

This case study was conducted at a leading manufacturer of microprocessor based electric motor control systems. The control units are designed on site in the UK, then manufactured and tested on one of two test rigs in Poland before being retested and configured on one of two test rigs in the UK ready for packaging and distribution.

The test rigs are known as GATE (General-Purpose Automatic Test Equipment) tests and involve testing both the hardware and software of the control unit. Prior to this investigation there had been an increasing number of control units failing the GATE tests, and it was found that approximately 20% of all GATE tests failed costing the company up to £800,000 per year in lost production time.

4.2.1 PVDT Implementation

After initial investigations the project team revealed that the number of faulty units returned to the engineers for further investigation following testing was relatively low. It

was recognized that the majority of fails were known as false fails; this is where a control unit fails the GATE test but then passes when retested.

The project team defined their project as finding the cause of false fails on the GATE test rigs in the UK. The team identified three separate Hardware False Fails that the GATE test can produce; High Armature Current Difference (ACD), Reverse Field Current Fails (RFC) and Battery Voltage Fails (BV). The most common cause of Fail type was ACD and so it was made the focus of the investigation.

In order to find the signature of variation a PVDT was performed on the two UK GATE test rigs. Five test times were selected across a day to take into account changes in shifts and breaks. At each of the five test times four units were tested three times, twice on GATE 1 and once on GATE 2.

The Gage R&R results came out at 65.8% when measuring ACD. This result was very concerning as it was well away from the QS9000 adequate guideline of 10% and suggested that the measurement system would need the immediate focus of the quality team. The capability study, shown in Figure 17, showed also that the process was neither capable nor centred.
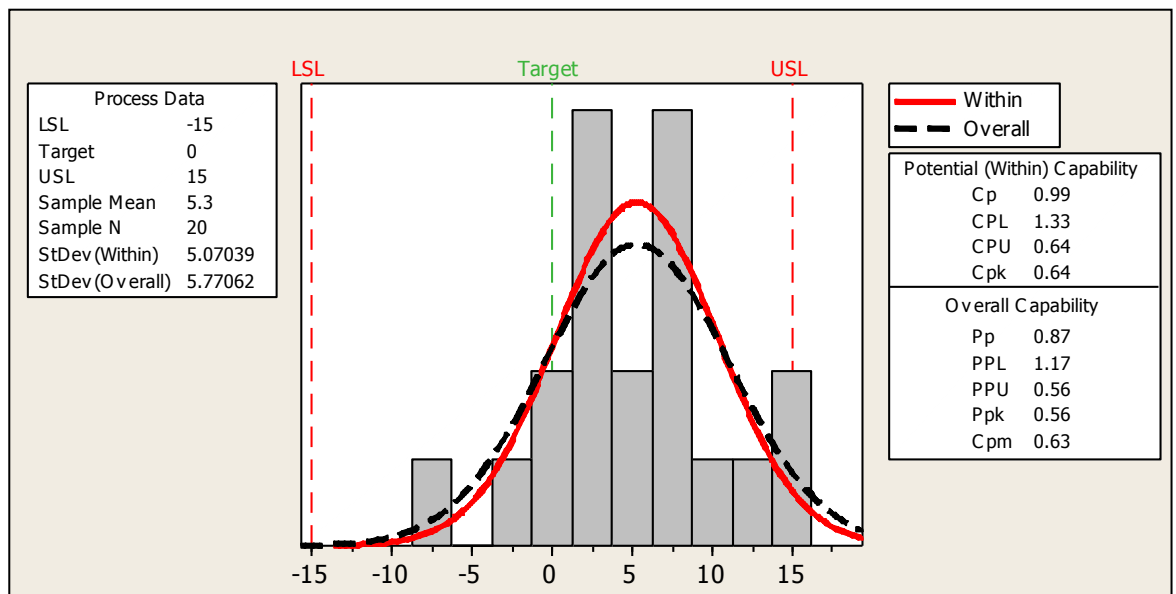
Figure 17 Process Capability Study for ACD, demonstrating the process is neither capable nor centred

The Red X^{SM} shown in the Shainin Multi-Vari in Figure 18 was clearly a within-piece problem which is supported by the large Gage R&R value. Figure 18 also shows that there is a lesser signature of variation or Pink X^{SM} across time-to-time and a Pale Pink X^{SM} piece-to-piece.
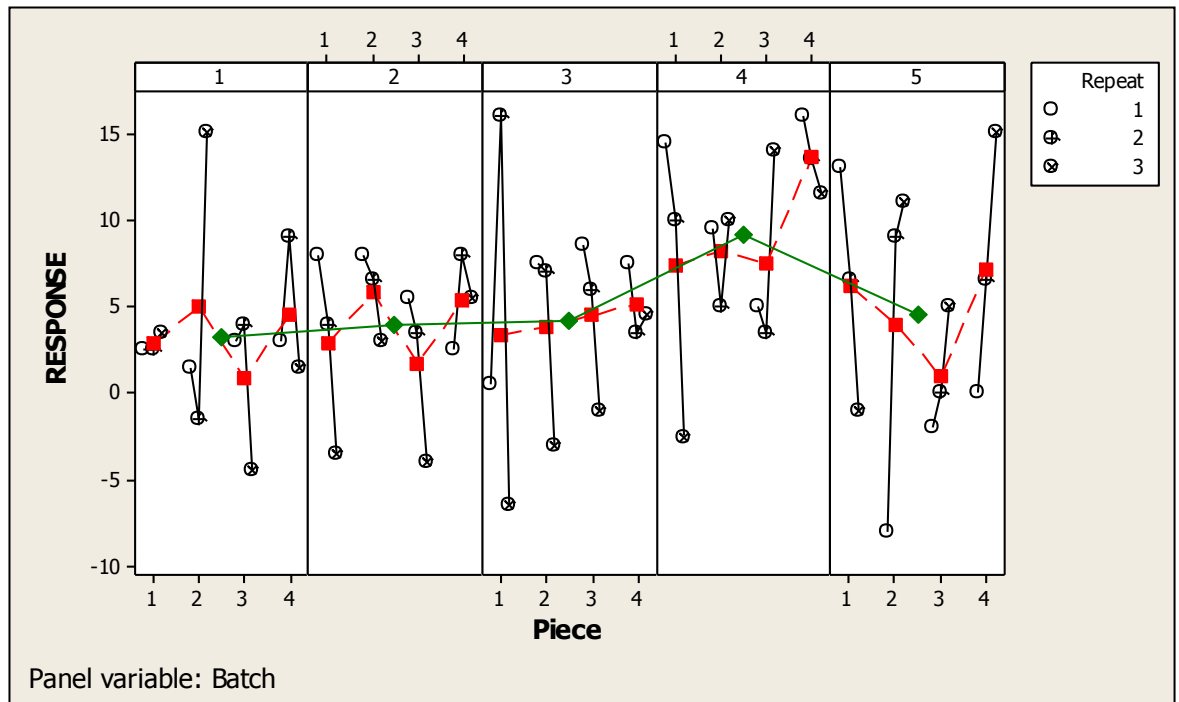
Figure 18 Shainin Multi-Vari Study for ACD, Showing a within-piece Red X^{SM}, a time-to-time Pink X^{SM} and a piece-to-piece Pale Pink X^{SM}

The within-piece variation was followed-up with an Isoplot^{SM} shown in Figure 19, demonstrating greater variation across GATE 1 test rig than across GATE 2.

The project team strongly suspected that the difference in variation across the test rigs was a result of a difference in armature current being used between the rigs. The follow-up to this was not possible within the 2 week project, but the project team were able to establish that there was a significant measurement system problem that was not a result of significant product variation or that the test process was affecting the result.
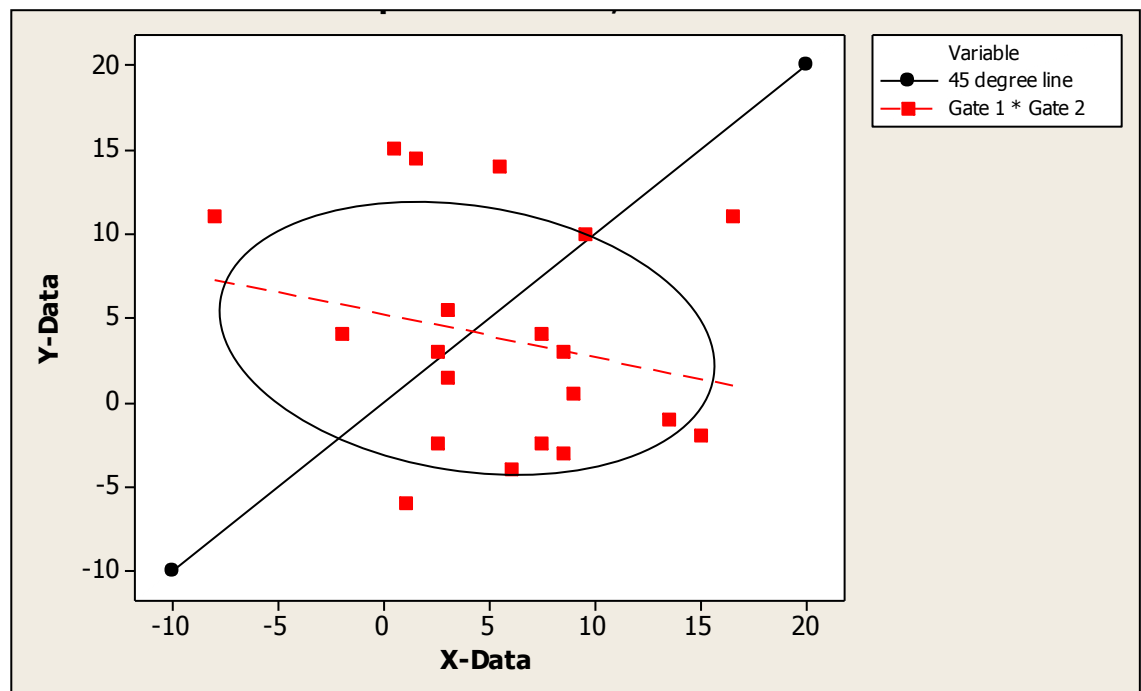
Figure 19 Isoplot^{SM} between Gates 1&2 for ACD, demonstrating greater variation across GATE 1

The time-to-time problem was followed up using a Cusum technique where the project team were able to link the changes in variation seen from time-to-time with changes in supply voltage from the Grid. At around 22:00 daily there was a spike in the supply voltage which was seen as an increase in the test results at the same time.

The piece-to-piece problem was left to last to solve as the resultant variation seen was smallest and as a result the project team were unable further narrow down the potential causes of this variation.

4.3 Case Three: Switched on Channels

The third case study was conducted at a major global electronics manufacturer. The company had a long standing quality problem on a low volume process producing microchannel plates for image intensifiers night sights. At the time of the improvement program their where 200 channel plates produced per year, production was planned to increase due to growing demand. Prior to this increase the yield from the process needed to improve from its current level of 25% which was resulting in the image intensifiers making a loss.

The first step was to Define the project. Table 3 shows the reasons for faults on the channel plates; from historical data it was found that 80% of faulty plates contained a SOC fault. It was decided that the focus of the process improvement program would be to resolve this issue.

Table 3 Description of channel plate faults

Fault	Description
Resistance	Electrical resistance between front & back of plate must lie between specified limits
Chicken Wire	Boundaries between multi-fibres (hexagonal pattern) show up (light or dark) when plate is working
Large Area Uniformity	Some areas of the plate have a different gain (brightness) when the plate is working
Background	With no input to the plate, there is an unacceptable amount of scintillation at the output (electrical noise)
Multi-Fibre Uniformity	Some single fibres within the multi-fibres have a different gain (brightness) which results in a recurring pattern over the plate
Switched-On Channels (SOC)	With no input, occasional channels have an output. This causes spots of illumination against a dark background (brightness depends on gain of SOC)
Black Spots	Inclusions of 'dead' channels give black spots against general illuminations when plate is working
White Spots	Occasional channels with extra high gain give white spots against general illumination when plate is working
Chips	Includes all mechanical damage to plates
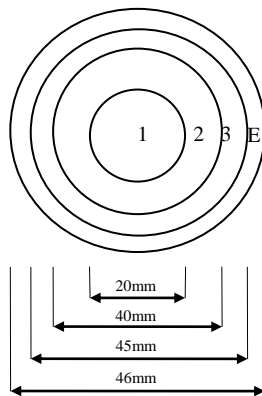
SOCs are classified in 4 brightness's:

B - bright

M - medium

D - dim

F - feint

Figure 20 Grading Grid for SOC on channel plates

Part of the project definition was determining how to classify the SOC problem. Figure 20 shows how SOC are graded for customer requirements. For a channel plate to pass it has to have fewer SOCs than in Table 4.

Table 4 Pass Grid for SOC

		Brightness			
		B	M	D	F
	1	0	0	1	2
	2	0	1	2	3
Zone	3	1	2	3	4
	E	2	3	4	5

A painful experience in the process improvement project was how to categorize the SOCs to produce useful information. Prior to the use of the PVDT the project team tried to establish links between SOC location, brightness and the process. Finally the team decided to just count the number of non-conformities on each sampled channel plate.

The Gage R&R results came out at 418% when measuring the number of SOC non-conformities. This result was extremely worrying, suggesting that there was a severe measurement system problem. The capability study applied a C-chart method to accommodate the attribute measures being recorded and is shown in Figure 21, demonstrating that the process is not capable.
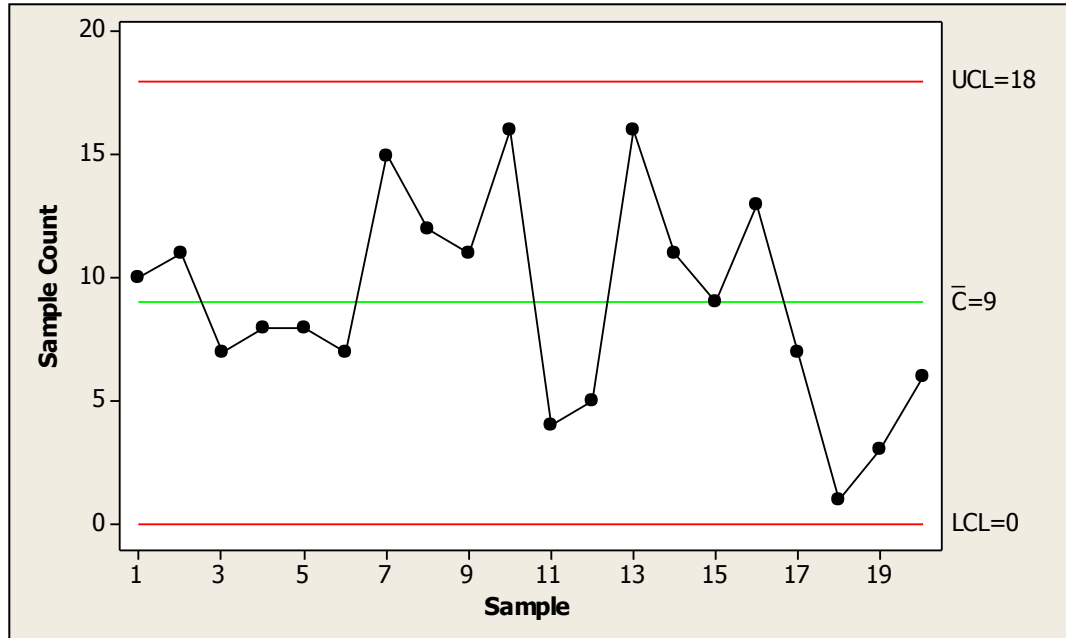
Figure 21 C-chart demonstrating the capability of the number of non-conforming SOC on channel plates
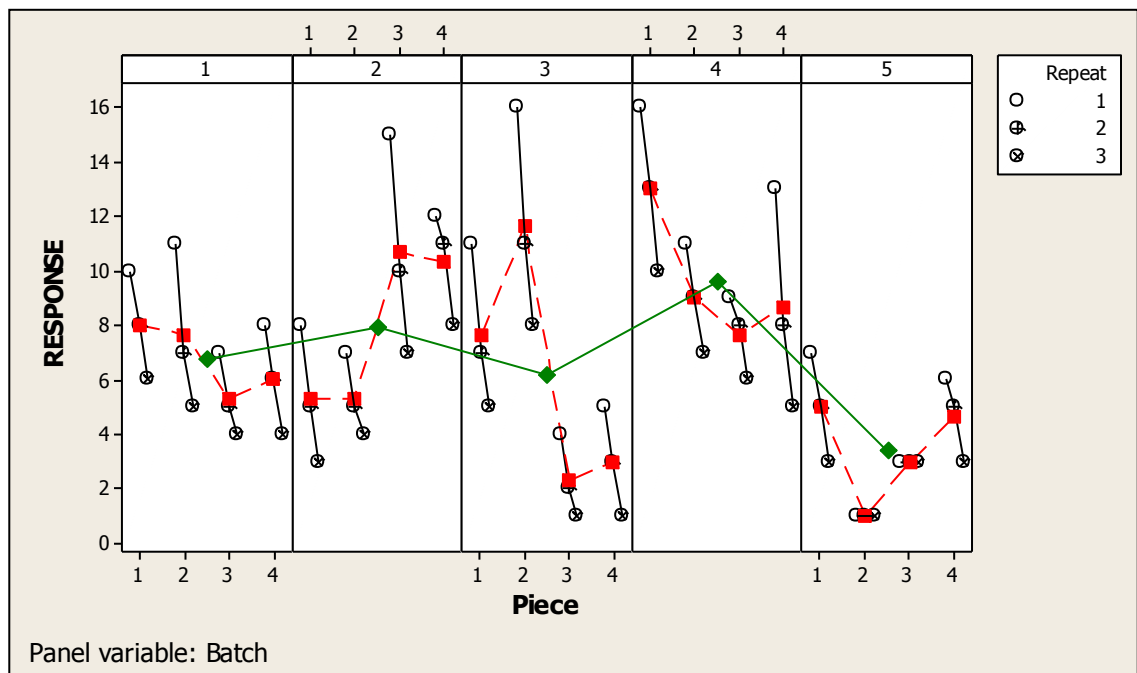
Figure 22 Multi-Vari showing a within-piece Red X[SM], a piece-to-piece Pink X[SM] and a time-to-time Pale Pink X[SM]

The Shainin Multi-Vari in Figure 22 shows a strong within-piece signature of variation which backs up the extremely large Gage R&R value for the SOC problem. This was immediately followed up with an Isoplot[SM] to further understand the measurement system issue. The Isoplot[SM] in Figure 23 shows the difference between Test Equipment 1 which is the current test rig and test equipment 2 which is an experimental test rig. Figure 23 shows

two "sausages" of points either side of the 45° line which are specific to whether the channel plate was tested on equipment 1 first or equipment 2 first. Thus this plot shows that irrespective of which equipment is used the 2nd test always displays fewer SOC non-conformities.
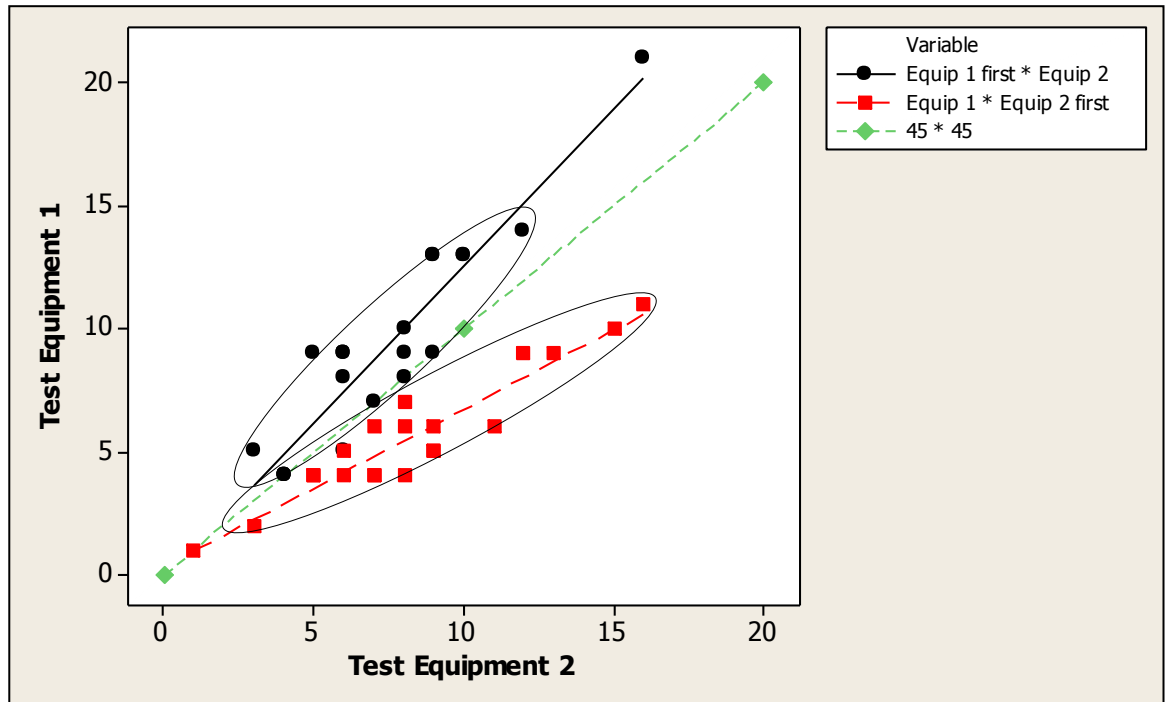
Figure 23 Isoplot[SM] for SOC measure system problem to compare variation between test equipments

The Red X[SM] within-piece problem was resolved by introducing an ageing process before the channel plates are tested, which lead to a significant increase in yield. This also suggests that the majority of previously disposed channel plates with SOC problems would have passed inspection if they had been retested.

5   TRW: Car Steering System Producer

5.1   Project Background

TRW Automotive is a global technology, manufacturing and service company that provide advanced technology, systems and services to customers worldwide. TRW is designing and producing world-class products for the automotive market in over 200 facilities.

TRW Houghton-le-Spring specialise in the manufacture of steering systems for customers, known as Original Equipment Manufacturers (OEM). One of the processes at the facility involves an automated laser welding process, which is used to weld Field-Effect Transistors (FET) to a stamping grid (Figure 24 shows the four FETs welded on the stamping grid and there numbering system left to right). This component is then used as part of TRW's power steering module.

Due to the number of parts been returned by OEMs, TRW's continuous improvement initiative is scrutinising the laser weld process. The current most accurate method of measuring the laser welds is using a microscope analysis of the weld penetration. This method involves a laborious process of polishing the edge of the weld and then photographing under a high magnification microscope. In order to obtain an accurate representation of the weld it can take up to a day of analysis per FET. The most practical test is a destructive pull test, but there are severe reservations over the consistency of these results.
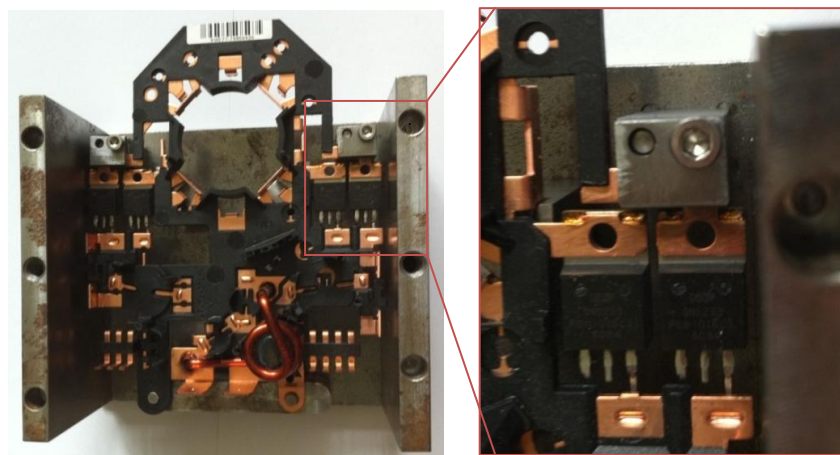
Figure 24 Photograph showing four FETs Laser welded to the Stamping Grid

In the fast-paced manufacturing environment in which TRW operate the accurate measurement method does not offer a practical way to measure the capability of the laser weld process. In order to overcome this problem TRW devised an experimental method to mechanically test the strength of the welds, using a non-destructive peel test method.

The test jig holds the stamping grid solid in a steel frame, as in Figure 24. A guide plate is then placed on top of the jig, to direct the probe on to one of the four FETs as in Figure 25. This jig can then be placed in a Lloyd LR5KPlus material testing machine, where a load can be applied to the probe, which is in turn applied to the FET causing the weld to peel and the deflection is recorded by the Lloyd.
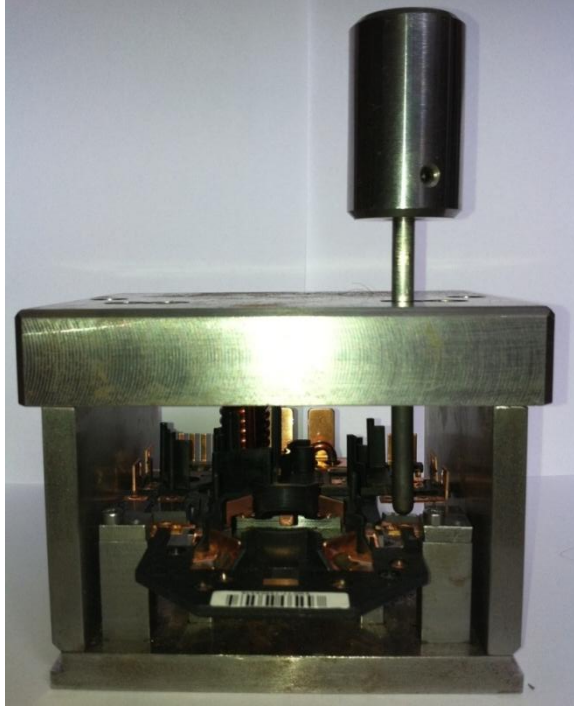
Figure 25 Complete Test Jig with Guide and Probe

Preliminary work had been carried out by two level 3 Durham engineers, to validate the design of the new test jig that would perform a non-destructive peel test. Their findings were reported in Hunter [22] and Waddell [23]. The principle findings from these investigations were:

- By measuring weld penetration depth it was shown that there was variation in the process between the FETs.
- A 5N limit load test offered the best opportunity for a non-destructive test, as the majority of samples started plastic deformation at around 7N.
- There were problems with the Lloyd test equipment at TRW cutting out up to 1N before load limits were reached, it was suspected but not proved that the 4mm/min test speed was too quick.

5.2 Project Definition

The key objective of the project is to validate the experimental 5N limit load test as a viable non-destructive test procedure. If possible the data should be used to gain further insight into the process driving the project forward from the Measure phase.

5.3   Initial Observations

During the set-up of the testing equipment a few potential issues were identified that could affect the test process. It will be shown that these issues were either resolved before the testing began or were considered in the test process.

Figure 26 (a) shows one of two pillars supporting the outer stamping grid arms, the one displayed is attached to FET 4. Whilst doing trial runs to check the Lloyd and jig set up, it was noticed that these pillars supporting stamping grid arms allowed the arms to flex when a load was applied to the FET. The resulting effect can be seen in Figure 26 (a) whereby the inner FET 3 is supported in a manner that causes the weld to peel when the load is applied to the FET. The outer FET 4 bends at the stamping grid arm thus reducing the load experienced by the weld. This problem was resolved prior to testing by adding a grub screw clamp to the top of the pillars; this can be seen in Figure 26 (b).
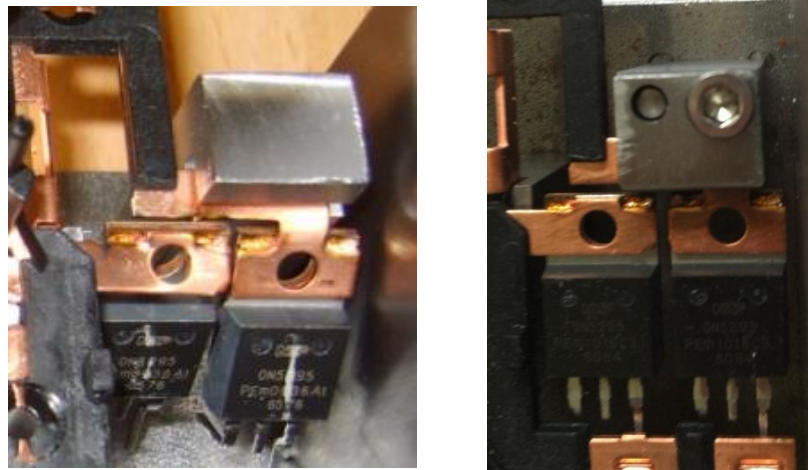
Figure 26 Showing (a) old pillar supporting outside FET 1 & 4 (b) new pillar with grub screw clamp

The Lloyd test equipment at TRW was configured with a 5kN load cell with an accuracy of ±0.5%. This could cause a potential measurement variation when applying loads of 5N, as it is a 1000th of the load cell maximum. Due to the cost involved with buying a new load cell it was not feasible acquire a new load cell with a smaller load limit. A 500N±0.5% load cell was available at Durham University. It was decided to use both load cells in order to assess the reproducibility of the measurement over the two load cells to demonstrate if these concerns were valid.

A final point not considered in previous investigations was that the FETs experienced a load and were already deflecting prior to the Lloyd test machine applying the 5N limit load. This can be seen in Figure 25 as the probe resting on the FET was applying a load of 2.3N and resulting in a deflection. It was therefore noted that the measure deflection by the

Lloyd test machine was the additional deflection created by the 5N limit load test and not the total deflection.

5.4   Test Method

To minimise the impact on current production at TRW a PVDT was implemented to determine the following:

- Gage R&R; to give a statistical assessment of variation present in the measurement system
- Isoplot[SM]; to give a graphical comparison of variation present between the two measurement machines
- Provisional Process Capability; to give a statistical assessment of variation present in the process
- Multi-Vari study; to give a graphical assessment of variation in a process

Due to the way samples are collected at TRW 24 samples were assessed, as opposed to the 20 samples normally used by the PVDT. Each sample was tested twice by the 500N Load cell at Durham then once by then 5kN Load cell at TRW; this should give information required for a Gage R&R. 12 samples were to be tested twice at Durham first then once at TRW, the remaining 12 were tested once at TRW first then twice at Durham; this should allow the Isoplot[SM] to demonstrate whether the test process is having an effect on the product.

5.5   Results

5.5.1   Correlation with Previous Results

It was seen during some tests, for example Figure 27, that there was some "noise" present before the FET began to deflect. It is suspected that this is created by friction between either the jig and the probe or the probe and the Load cell when the jig is not directly beneath the Load cell. This "noise" was not present on all tests and demonstrates a potential source of operator error. This potential operator error was not considered in the structure of the PVDT as it was initially identified that the measurement equipment would be the biggest source of variation but this operator error will need to be quantified in the future.

Part of this project was to compare results recorded in Hunter [22] and Waddell [23] with the results captured in the PVDT. This was not possible as the deflections obtained by Hunter and Waddell were in the order of 1.5 mm, the results captured during this series of testing were in the order 0.2-0.6 mm. It is suspected that this can be explained by the initial "noise" seen in Figure 27. The results captured for the PVDT where exported from the Lloyd test machine software to Microsoft Excel in order to calculate the deflection as

in Figure 27. It is believed in Hunter and Waddell's experiments a preload was set on the Lloyd and the final deflection was recorded, although this method offers speed, it is likely that the noise seen would be enough to trigger the preload and thus the measurements taken would be much larger and not accurate.
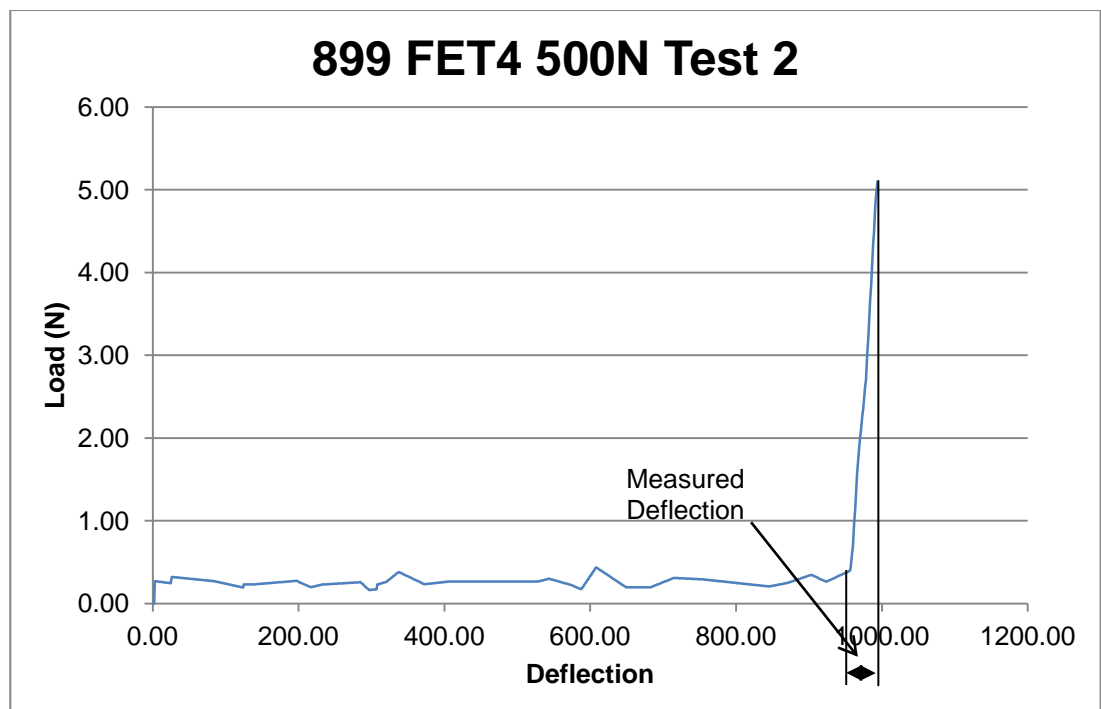
Figure 27 Example response from a 5N limit load test

5.5.2   Isoplot^{SM}
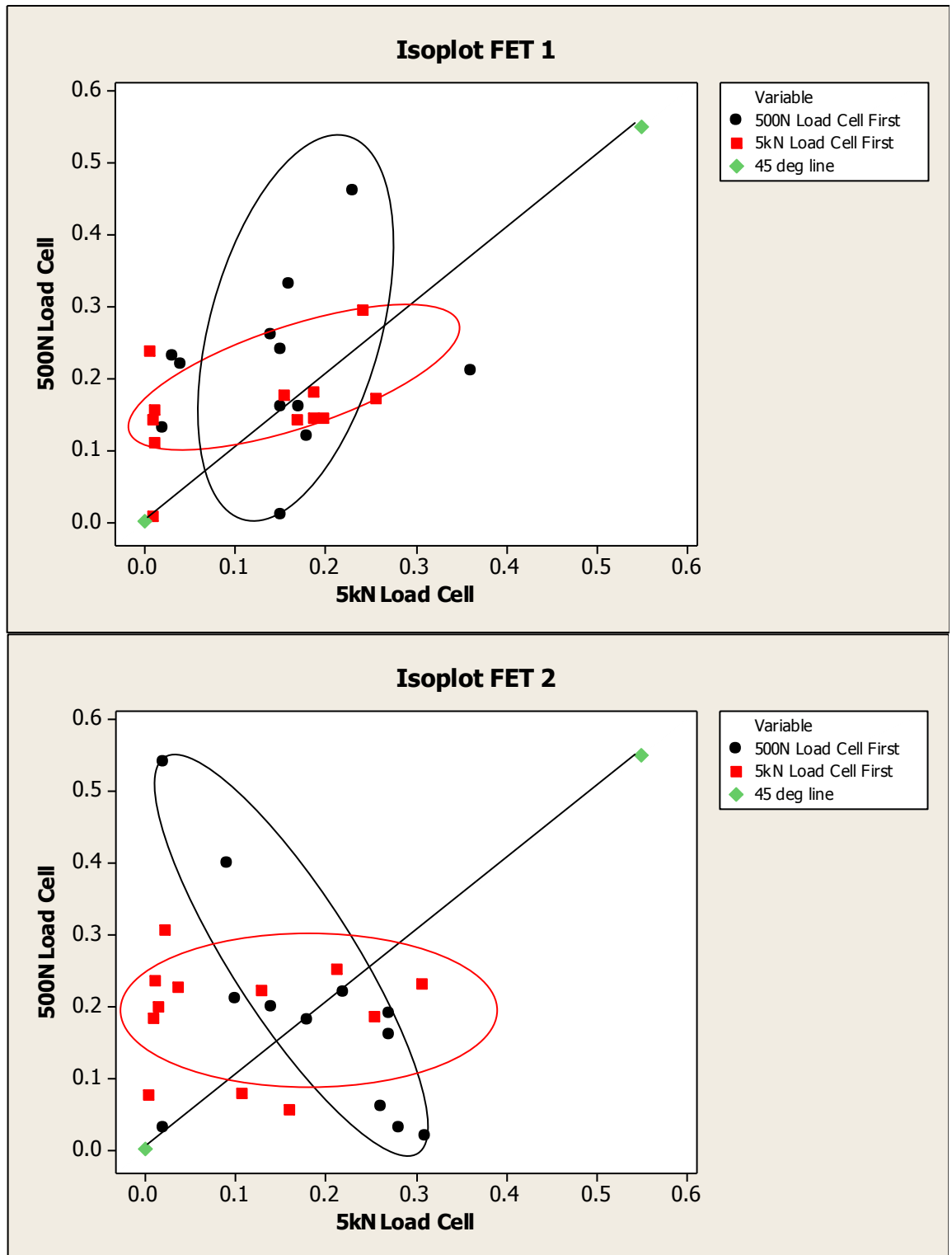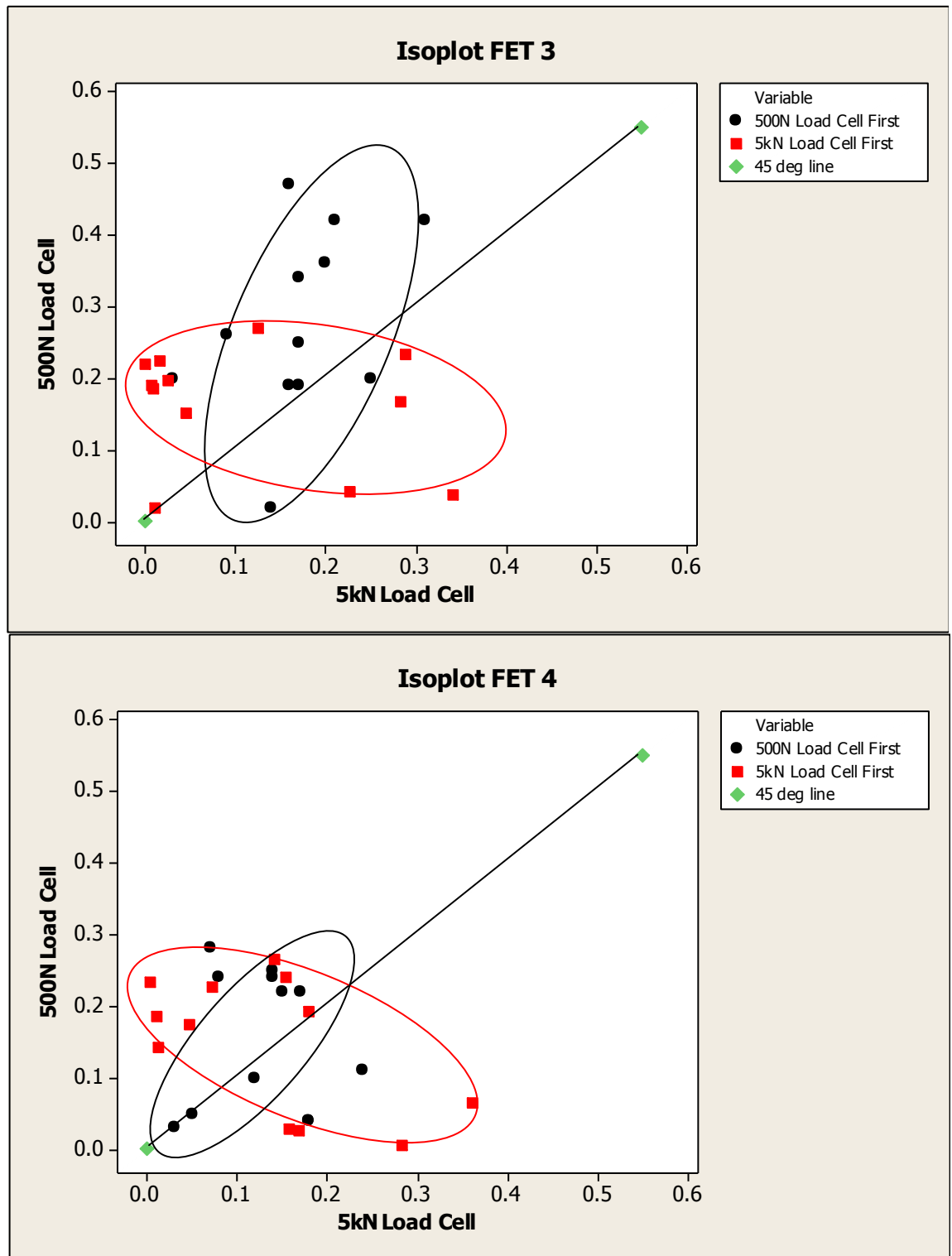
Figure 28 Isoplot analysis of results for FETs 1,2,3 & 4

The four Isoplot[SM] results from testing the FETs are shown in Figure 28. They demonstrate that the variation present due to the measurement equipment is large compared to the variation present due to the product variation, i.e. the spread of results away from the 45° line is greater than the spread of results along the 45° line. There is also a reflection of the two groups around the 45° line. This would suggest that the process of testing the welds was in fact affecting the characteristics of the welds; it is therefore a destructive test.

5.5.3 Multi-Vari study
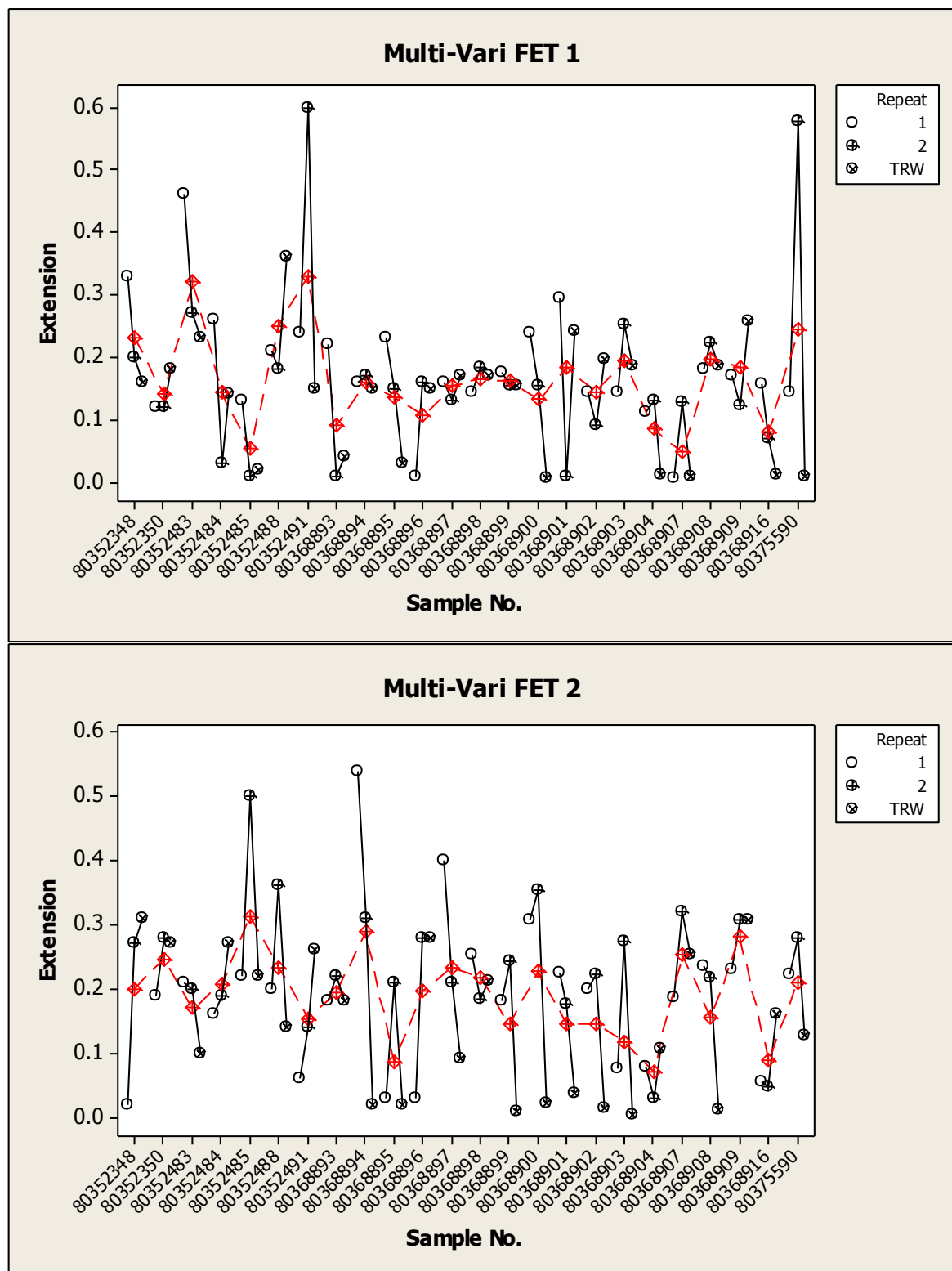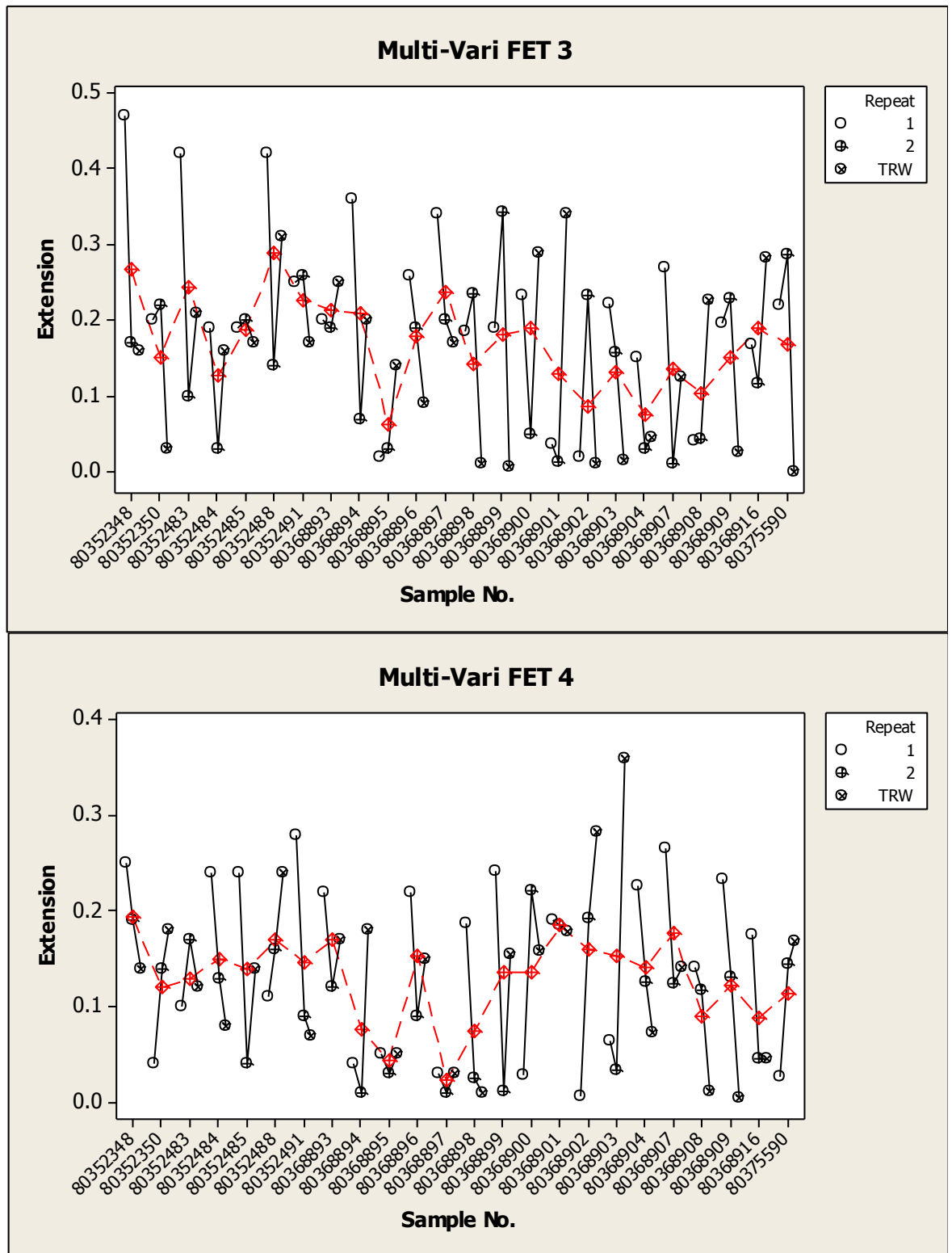
Figure 29 Multi-Vari Studies for FETs 1, 2, 3 & 4

The Multi-Vari studies in Figure 29 demonstrate that the within piece family of variation contains the Red X[SM]. This can be explained as a result of measurement variation with Isoplot[SM] analysis. These studies also show a Pink X[SM] related to the piece-to-piece family of variation. This is consistent with the weld penetration depth measurements taken by Hunter and Waddell that suggested there was process variation between FETs. However

the size of the piece-to-piece variation may be masked or inflated due to the large within-piece variation.

5.5.4 *Visual Observations*

It was noted that after testing was complete, many of the FETs were deformed. On three occasions this can be attributed to operator error, as the probe was dropped on the FETs by accident. This however does not account for all the deformed FETs which again support the Isoplot^{SM} analysis that the measurement process is having a detrimental effect on the characteristic of the welds.

From the testing it was not possible to carry out statistical tests against tolerance. Due to the large measurement variation demonstrated by the Isoplot^{SM} it was not possible to determine realistic tolerances. Statistical tests could be calculated against part-to-part variation but these results will not add any real information to the analysis at this stage given the large measurement variation demonstrated likely to skew the results.

5.6   Disscusion

Given that this test procedure is destructive; in the sense of changing the characteristic of the weld as a result of testing. Counter to the initial belief that this method was non-destructive. There are two possible ways of driving the project forward:

- Develop the method further to gain a non-destructive method; it would be beneficial to mount the probe directly onto the Load cell in future testing. This would eliminate friction between the probe and the Load cell. It would also allow the total deflection of the FETs to be measured and reduce the total load experienced by the weld from 7.3N to 5N. This reduction in load may be enough to stop the test affecting the weld characteristics.
- Develop a second destructive experimental test method; a second experimental method which test the weld in sheer to yield. This method was initially developed in parallel to the peel test but was put to one side as the peel test was theoretically more attractive. The destructive test is more difficult to validate and it would be a more expensive ongoing measure due to the cost of material waste.

5.7   Conclusions

This project clearly demonstrated the effectiveness of the PVDT. From 24 samples it was proven that the measurement system was inadequate as a method. It can also be seen that if the measurement system was valid, it would have been possible to produce a Gage R&R and Process Capability statistics which fulfil the final measure and early analyze of the Six Sigma process improvement cycle.

It also produced a Multi-Vari study which confirmed the focus on the measurement system was the correct action, as the signature of variation was clearly within-piece. It is possible to see that if the measurement system was valid the Multi-Vari would assist the process improvement initiative home in on the root cause of the problem. It does this without capturing any additional information than what is required for a Gage R&R and in an objective manner.

It is clear to see the benefit of a graphical analysis of the measurement system utilizing the Isoplot^{SM}. Supporting a Gage R&R study with a graphical analysis gives a clear indication where a measurement problem with a large Gage R&R result stems from. In the case of TRW where tolerances were not obtainable it is of even greater importance to have a graphical test.

After presenting these findings to the process engineers at TRW, they were able to objectively decide to end their follow up of this measurement method. This decision was made largely based on the Isoplot^{SM} analysis, as the large measurement variation would possibly require the purchase of a new load cell that is more sensitive to the low loads applied by the Lloyd test machine. The issue of the test affecting the characteristic of the weld was a second consideration as this would need a reduction of the limit load being applied which would again point towards the need for a more sensitive load cell. This ultimately may not prove effective.

One final point is that the lead engineer on the project from TRW estimated that the efficiency of the PVDT had saved TRW around £10,000 compared to their usual routes of validating an experimental measurement system.

6 PVDT Follow-Up Investigations

As discussed in previous chapters, the Shainin Multi-Vari study can help decide which Shainin or Six Sigma technique would be most appropriate to further investigation. Components Search and BOB vs WOW are two of the techniques suggested and are both Shainin techniques. As with many Shainin techniques they are based on non-parametric statistical tests as they are easier to use and teach. This is an advantage in an engineering environment, where statistics are often put to one side over judgement. Statisticians will quite rightly argue that that this ease of use will be at the expense of statistical power. However the practical power of test, as defined by Churchill Eisenhart [24], is a product of the mathematical power by the probability that the procedure will be used.

6.1 Best of the Best vs Worst of the Worst (BOB vs WOW)

This technique, also known as paired comparisons, collects a sample of at least four of the very best (BOB) products compared to the designed target and the same number of the very worst units (WOW) products. It then compares parameters or quality characteristics and then ranks them to find if any of these parameters have an effect on whether the unit is good or bad. By ranking the units we can use a Tukey end count test [24] to attribute statistical significance to the result. Using the example size of four of the best units and four of the worst units we can gain a confidence in any findings of up to 95%. With a larger sample size of six of the best units and six of the worst units we can gain a confidence of up to 99%.

This is a very simple yet powerful technique to further reduce the search for the root cause of variation. It also does not rely on experimentation with process settings which means it can be conducted without interrupting production, it merely looks at samples from ongoing production.

6.2 Components Search

In low volume high value manufacture new statistical methods to find objectively the root cause of defects and variation of critical-to-quality (CTQ) characteristics in a product are constantly being pursued. This Section will outline and provide case study results of a novel but often overlooked technique known as Shainin's Component Search. It is a process improvement technique which is particularly useful as a follow up to a PVDT investigation when there is an assembly which can be disassembled and reassembled.

The Component Search procedure was pioneered by the American quality guru Dorian Shainin in the early 1960s whilst working with management consultant firm Rath & Strong Inc on problems with assemblies of a great number of parts [8]. It is used to compare good

and bad assemblies when two or more assemblies are available that can be disassembled and reassembled without effecting the CTQ [8][12][10].

The component search procedure as described in Shainin and Shainin [8] and Bhote [12] consists of four stages. Stage one, known as the ballpark stage, tries to determine if the root causes of defects are among the components being considered and also assures the repeatability of the assembly process. Stage two, the elimination stage, seeks to eliminate unimportant components from the search. Stage three, the capping run, verifies that the important components have been identified. Stage four, factorial analysis, demonstrates the size of important components and if there is any interaction effects between them. It is conducted with a sample size of 2 assemblies, with one assembly being a worst of the worst (WOW) unit and the other being a best of the best (BOB) unit. According to Bhote [12] these two assemblies should be at the "extremes of the distribution".

This means the method as describe is easy to apply when the CTQ's have a single sided specification, i.e. the bad products are all one side of the process target. When there is a two sided specification, i.e. bad products either side of the process target using only two samples will either break the principle of comparing BOBs and WOWs or it will not consider the whole distribution. The next section of this paper will outline the procedure described in Shainin and Shainin [8] and Bhote [12], but modifications to this method will also be proposed using a sample size of 3 to make it applicable to the two sided specification case without breaking the principles of comparing BOBs and WOWs and also using units at the "extremes of the distribution" thus considering the entire distribution.

6.2.1   Components Search Procedure

.6.2.1.1   Stage 1: Ball Park

Select two sample assemblies from production, one unit being a Best of the Best (BOB) unit and one being a Worst of the Worst (WOW), with their measured CTQ as far apart as possible. This is the traditional approach for a quality problem with a single sided specification [12] (i.e. a CTQ with either an Upper Specification Limit (USL) or a Lower Specification Limit (LSL)). It is proposed that when there is a double sided specification (i.e. a CTQ with both USL and LSL) that three sample assemblies are used for the analysis. One of these assemblies should have a CTQ as close to the process target as possible, thus is the BOB unit. The other two assemblies should be WOW units with one from the upper extreme of the process variation and other from the lower extreme of the process variation.

At this stage, irrespective of whether two or three assemblies have been sampled, the sample assemblies should be disassembled and reassembled twice and there CTQ

measured after each reassembly. These results are then used to test the validity statistically of performing the components search method on the assemblies

The first statistical test of repeatable difference is ranking the measured CTQ's from the disassembly/reassembly experiments. There should be no overlap in the results between BOB and WOW assemblies to pass this test. The test is based on the non-parametric Tukey End Count test, where no overlap in CTQ will give an end count of 6 units which in turn will give 90% statistical confidence that the BOB and WOW assemblies have different means [24]. Table 5 gives an example of a BOB assembly with a measured CTQ in the range 96-100 and WOW assembly with a CTQ in the range 68-72, after the disassembly/reassemble process. This example would pass the test of significance as there is no overlap in the measured CTQ's.

Table 5 Example disassembly/reassembly which contains no overlap in measured CTQ

	BOB	WOW
Initial	100	70
First Reassembly	96	68
Second Reassembly	98	72

Table 6 gives an example of a BOB assembly with a measured CTQ in the range 81-100 and WOW assembly with a CTQ in the range 70-85, after the disassembly/reassemble process. This example would fail the test of significance as there is an overlap in the measured CTQ's. In this case the assembly process itself is having an effect on the variation seen and needs to be investigated.

Table 6 Example disassembly/reassembly with overlap in measured CTQ

	BOB	WOW
Initial	100	70
First Reassembly	81	76
Second Reassembly	93	85

The second test of repeatable difference is to ensure the D:$\bar{d}$ ratio must be greater than or equal to 5. Were D is difference between medians of the BOB unit and WOW unit for a single sided specification or the difference between the extreme WOW units for a double sided specification. $\bar{d}$ is the average of the ranges of the three BOB measurements and three WOW measurements (there will be an extra set of WOW ranges when 3 samples are considered). If the D: $\bar{d}$ is less than 5 this indicates the components search has failed and indicates a problem with the assembly process, as the CTQ's do not remain constant.

Using the data from the Ball Park stage decision limits are calculated for the elimination stage. If the components search is being conducted with two assemblies this can be calculated using formula (1); which uses the median of the disassembly/reassembly results and $\bar{d}$. The statistical constant 2.776 is derived from a two-tailed t-distribution with 95 percent confidence and 4 degrees of freedom (as 3+3-2) and 1.81 is statistical constant, d_2^*(for 2 samples with sample size 3)[25][12].

$$Decision\ Limits\ Single\ Sided\ = median \pm 2.776 \times \frac{\bar{d}}{1.81} \qquad (1)$$

If the components search is being conducted with three assemblies this can be calculated using formula (2); which is an adaption of formula (1). The statistical constant 2.447 is derived again from the two-tailed t-distribution with 95 percent confidence but with 6 degrees of freedom (as 3+3+3-3) and 1.77 is statistical constant, d_2^* for 3 samples with sample size 3..

$$Decision\ Limits\ double\ Sided\ = median \pm 2.447 \times \frac{\bar{d}}{1.77} \qquad (2)$$

.6.2.1.2   Stage 2: Elimination

Rank the components in descending order of likely significance; this could be done through engineering judgement or in the worst case arbitrarily. Although this is a subjective method, it will only affect the speed of the component search [10] and not the likelihood of finding the root cause of the quality problem.

Switch the top-ranked component from the BOB to the WOW and vice a versa. Measure and record the new CTQ's. For the case where three samples are required the same process applies except the BOB component is swapped with associated WOW component from a WOW assembly at one side of the specification and then swapped with the WOW assembly at the other side of the specification given four new measured CTQ's.

There are three possible outcomes:

1. No change in observed CTQ; swap the tested component back to its original assembly and move on to the next component.
2. If there is a small change in observed CTQ in one or more assemblies this component is important but not the only component responsible for the variation seen. Swap the tested component back and move on to the next component.
3. If there is a complete switch in measured CTQs; i.e. a BOB assembly becomes a WOW assembly and the WOW assembly becomes a BOB assembly, this component is responsible for the variation seen and the search is stopped. For the two sided specification case the complete switch must appear on both sides of the test.

.6.2.1.3 Stage 3: Capping Run

Once it is suspected all components identified as important to variation seen. A capping run is performed; whereby all important components from the BOB assembly are swapped into the WOW assembly and all important components from the WOW assembly are swapped into the BOB assembly to check if the measurements of the respective CTQs have also completely swapped. This test is run twice in the two sided specification case. The capping run confirms whether or not all the important components have been found. If the important components have been identified the analysis moves on to the next stage the factorial analysis and if they haven't the analysis moves back to the elimination stage and components continue to be swapped till all important components are found.

.6.2.1.4 Stage 4: Factorial Analysis

A full factorial analysis is drawn up using all the collected results from stages 1 and 2 to assess the size and relationship of the main effects and interaction effects.

6.2.2 Simulation Case Study

The application of Components Search is now illustrated using a simulation known as the PIM Game which is used to demonstrate Process Improvement Methods (PIM). The game is based around production of an assembly which is made up of 6 components. The values for the 6 components, known as A, B, C, D, E, and Substrate, are generated randomly with in specific ranges to represent variation in components. They also represent a measureable critical parameter; this could be resistance, diameter, etc. Participants build up assemblies in the PIM game over 4 workstations; it is then measured and tested at separate stations. From the test station the assembly will have numerical and Pass/Fail result with a double sided specification with a range of 20-180, whereby a score of 100 is best. The further away from 100 the product scores the worse the product is. With an USL set at 130 and a LSL set at 70, i.e. a score in the range of 70-130 means the fault has "passed" the quality check.

Traditionally a Fractional Factorial followed by a Full Factorial was used to solve this problem which required 16 "online" experiments. Using Components Search we were able to demonstrate by obtaining 3 assemblies, 1 good assembly and 2 at the extremes, the main causes could be determined "offline".

.6.2.2.1 Stage 1

Table 7 shows the 3 assemblies selected to run the Components Search Experiment. They were selected as the very best unit sampled and two of the worst units sampled, out of specification either side of the centre.

Table 7 Selected Assemblies for Components Search Procedure

	A	B	C	D	E	Substrate	Sample Results	Reassembly 1	Reassembly 2	Median	Range
Bad	6	7	1	0	3	0	B	176	176	175	176	1
Good	9	8	11	10	3	1	R	100	98	98	98	2
Bad	6	13	18	9	3	2	Y	36	35	36	36	1

Table 7 demonstrates that the samples pass the first significance test for repeatable difference that none of three measured CTQ results for each unit overlap, i.e. the three repeated measures for the WOW assembly which CTQ is below the LSL do not overlap with the three repeated measures of the BOB assembly. Similarly there is no overlap between the WOW assembly with a CTQ above the USL and the BOB assembly.

We can now show $D = 176 - 36 = 140$ and $\bar{d} = \frac{1+2+1}{3} = 1.333$ which means the ratio; $D : \bar{d} > 105 : 1$. These results pass the second significance test for repeatable difference which means the source of variation is in the components and not a result of assembly techniques.

The decision limits were calculated, in equations (3), (4) and (5), using the information for the assembly medians from Table 7, the calculated $\bar{d}$ value and the decision limit formula for the a double sided specification proposed in formula (2).

$$Decision\ Limit\ High = 176 \pm 2.447 \times \frac{1.333}{1.77} = 174.15\ and\ 177.85 \qquad (3)$$

$$Decision\ Limits\ Good = 98 \pm 2.447 \times \frac{1.333}{1.77} = 96.15\ and\ 99.85 \qquad (4)$$

$$Decision\ Limits\ Low = 36 \pm 2.447 \times \frac{1.333}{1.77} = 34.15\ and\ 37.85 \qquad (5)$$

.6.2.2.2   Stage 2 and 3

The Components Search Experiments were run plotting the results in Figure 30. Figure 30(a) plots the results between the WOW assembly with a CTQ above the USL and the BOB assembly. It demonstrates that components A, C and D are important factors in terms of affecting the resultant CTQ of an assembly. Figure 30(b) plots between the WOW assembly with a CTQ below the LSL and the BOB assembly. It confirms clearly A and C are important factors and also just shows D falling outside the BOB assembly decision limit.

As a result of the components search analysis, capping runs are performed swapping components A, C and D from the WOW assemblies to the BOB assembly to test if all important factors have been captured. The results of the capping run are also shown in Figure 30(a)(b) with these plots clearly demonstrating that components A, C and D are the important factors and account for all the variation seen in the assemblies measured CTQ.

(a)

(b)
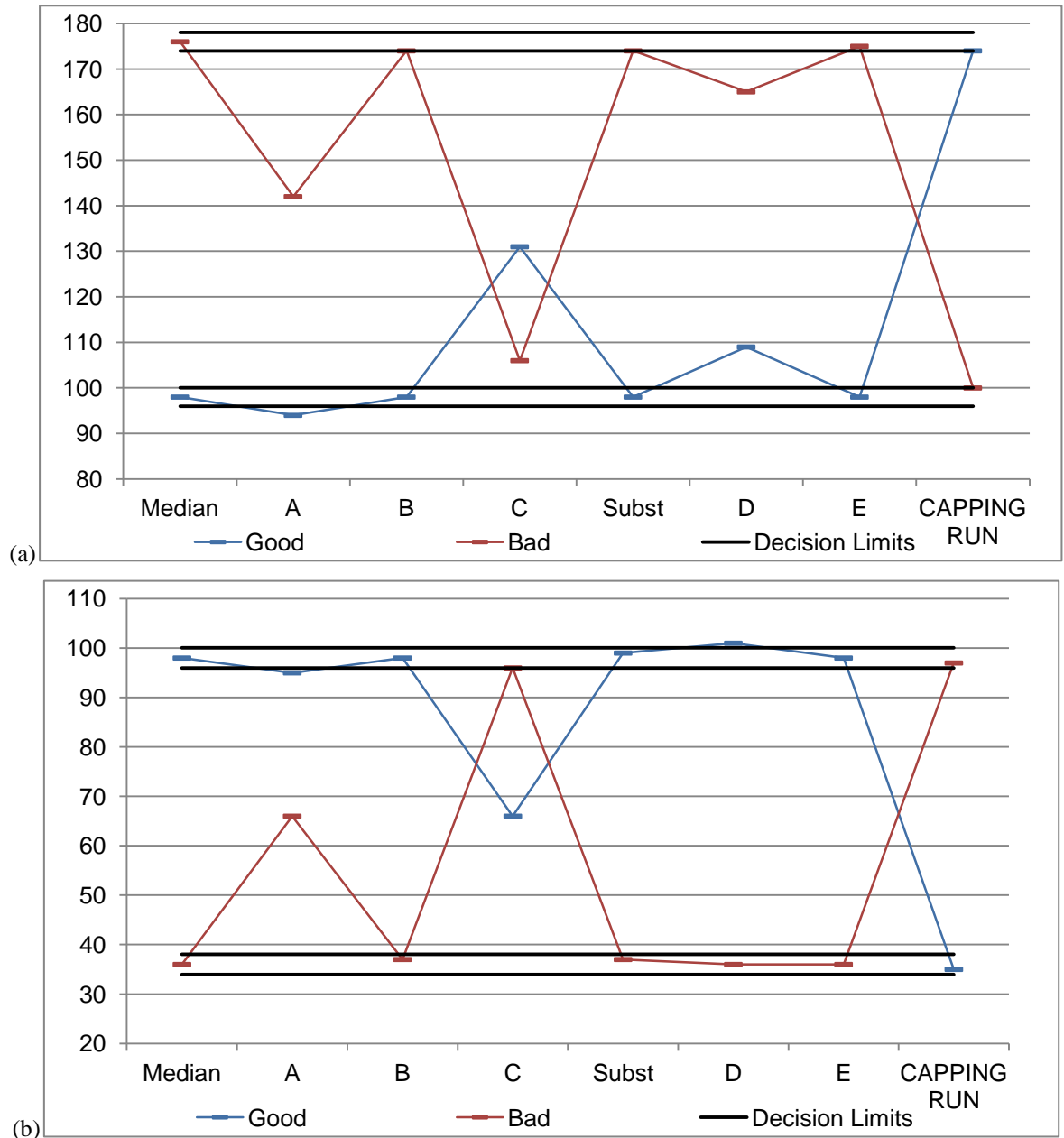
Figure 30 Graphical Plot with Decision Limits of Components Search between (a) Good and Bad High (b) Good and Bad Low

.6.2.2.3 *Stage 4*

The final stage of the procedure is the full factorial analysis to quantify the size and relationship of the main and interaction effects. As the elimination and capping run stages indicate that factors A, C and D are the ones important to the measured CTQ of the assembly the factorial analysis can be limited to these factors.
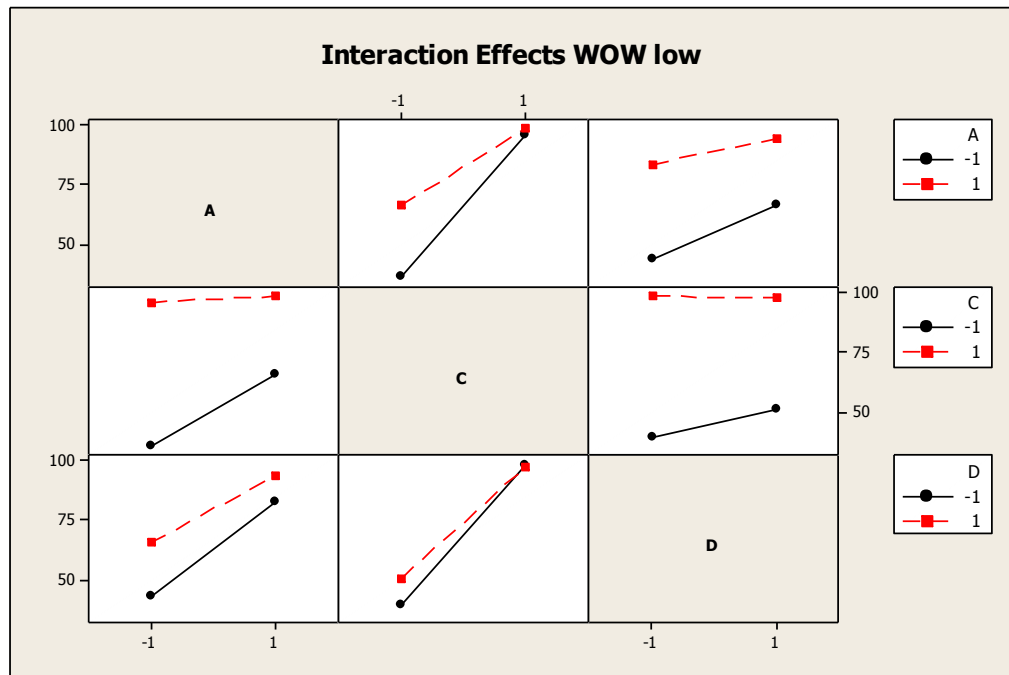
Figure 31 Factorial Analysis of Factors A, C and D for BOB v WOW low

Figure 31 demonstrates the main and interaction effects between the BOB assembly and the WOW assembly on the low side of the specification. This figure confirms the capping run results that A, C and D are important. It shows clearly there is an interaction between factors A and C; as two interaction plots between these two factors show lines which converge and are non-parallel. The plots for A and C also confirm that these components from the BOB unit are the best setting for the CTQ of the assembly. The plots for the D component do not show an interaction with other components as the lines on these plots are largely parallel and the size of the effect of this component is small compared with the effects of A and C. It does show that the CTQ is closer to the process target of 100 when D is at the BOB setting.

Figure 32 shows the main and interaction effects between the BOB assembly and the WOW assembly on the high side of the specification. This figure again confirms the capping run results that A, C and D are important. It shows clearly there is an interaction between factors A and C; as two interaction plots between these two factors show lines which converge and are non-parallel. The plots for A and C also confirm that these components from the BOB unit are the best setting for the CTQ of the assembly. The plots for the D component do not show an interaction with other components as the lines on these plots are largely parallel and the size of the effect of this component is small compared with the effects of A and C. It does show that the CTQ is closer to the process target of 100 when D is at the BOB setting.
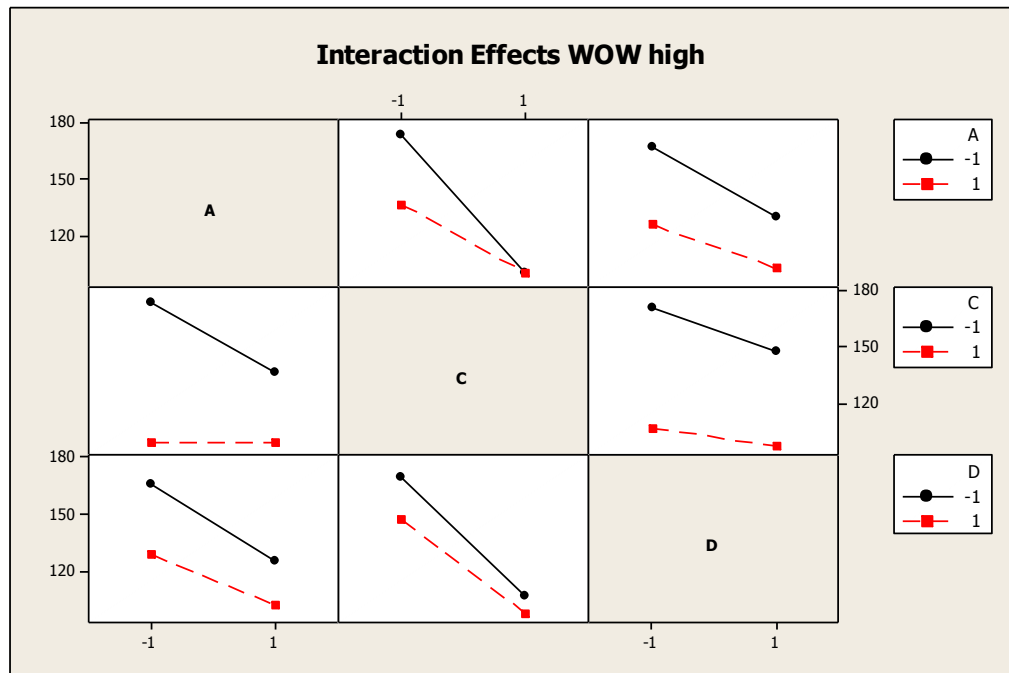
Figure 32 Factorial Analysis of Factors A, C and D for BOB v WOW high

6.3   Cusum

The Cumulative Sum or Cusum technique is useful when you are attempting to identify what triggers a time-to-time Red X source of variation. In order to achieve this, groups of 4 samples are collected and tested. If the time-to-time variation is across a batch process these groups of 4 samples should be taken from consecutive batches until a change in the process is identified. If the time-to-time variation is across a shift, day or week in a flow line, 4 consecutive samples should be collected at consistent intervals such that a change in the process can be detected.

Once samples have been collected, with the suspect CTQ measured and recorded, the difference between the process target and the measured CTQ should be calculated for all samples. For example if a process has a target of 100 units and a sample has a CTQ of 120 units the difference is +20. However if the CTQ is 75 units the difference is -25.

The cumulative sum of these differences should then be calculated and plotted. Figure 33 is an example of a Cusum from a batch process. The batches from which the samples have been taken are numbered and it clearly shows there is a change in the process between batches 2 & 3, 4 & 5, 6 & 7 and 8 & 9. The next stage is to look for changes in the inputs to the process at these times that result in changes in the CTQ.
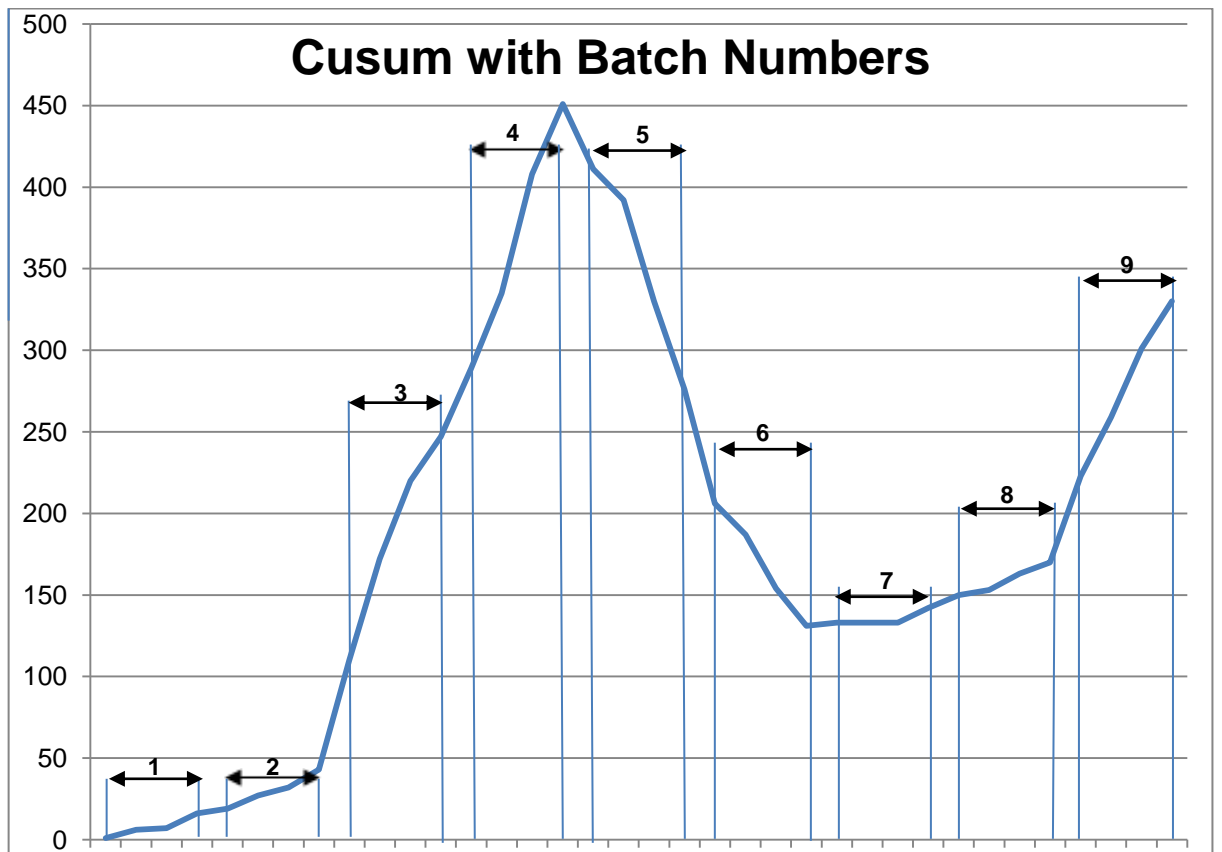
Figure 33 Example Cusum from a batch process

7   Process Improvement Methodology Game

The Process Improvement Methodology (PIM) game is a manufacturing process simulation used in a class room environment to demonstrate practically a range of Six Sigma and Shainin process improvement tools within the Durham Hybrid DMAIC framework outlined in previous chapters.

7.1   Context

There is a need for practical demonstrations of process improvement methods in the classroom. Using hands on demonstrations aids the participating students or professionals understanding of the demonstrated material. It allows participating students to explore the application, potential and limitations of new techniques, which builds confidence for them to be successfully implemented beyond the classroom. Developing this confidence and understanding can be difficult to achieve for all students through lectures alone. The Chinese philosopher Confucius stated "I hear I forget, I see I remember, I do I understand", it is with this philosophy in mind that new teaching experiments, simulations and tools are developed.

There are examples in the literature of practical experiments to demonstrate both Shainin and Six Sigma process improvement tools. A paper helicopter experiment is described in Anthony [25], eloquently illustrating how Shainin's Variable Search procedure can be applied to the production process of paper helicopters to identify the key variables which affect flight time. Bhote [12] suggests a range of examples and workshop exercises, such as a card game example to make evident the power of the "homing in" concept behind many Shainin tools. The workshop exercises however tend to consist of data taken directly from a case study and therefore lack the full interaction of the student completing the full cycle of deciding how to organise data collection in order to implement the specific tool. A bubble mixture experiment is proposed by Steiner [26], were participating students need to find the optimum ratios and types of water, soap and glycerine in order to produce a mixture which makes the greatest number of bubbles when a fixed amount of air is passed through the mixture. In order to complete this task DOE techniques which are commonly used in Six Sigma methodology, such as fractional factorials are employed, as well as optimization tools including Taguchi and Response Surface modelling and more advanced statistical modelling techniques. A more complete collection of hands-on examples of Six Sigma techniques is compiled in Chen [27]. Scripts for demonstrators are given to run practical experiments in a classroom illustrating tools and techniques used within the DMAIC structure of Six Sigma. However the methods go from the subjective such as brainstorming to more complex DOE. As discussed in earlier chapters on Six Sigma

literature there are little tools offered in the analysis to aid objective analysis without resorting to complex statistical tools.

Although these demonstrations offer clear examples of how to apply specific process improvement techniques, they do not provide a continuous environment were tools are demonstrated in the context of process improvement framework such as Six Sigma. Steiner [28] introduces Watfactory, which is a manufacturing process environment where students can investigate quality issues using their knowledge of process improvement. It is accessed via the Watfactory website and allows students freedom to request samples from the process and apply process improvement methods to determine the root cause of quality problems in the simulated factory. It also adds realism by applying financial constraints to the students' project. Although this simulation is the most complete teaching tool in terms of allowing a student to go through stages of process improvement cycles, it is however an excellent consolidation tool for students who have already been exposed process improvement methods and not a method of interactive demonstrating within the classroom. This could be considered as a follow up to the PIM game rather than an alternative method.

7.2   PIM game

The PIM game is a unique teaching tool used within workshop sessions to demonstrate a blend of Six Sigma and Shainin quality improvement tools. The rationale behind using a blend of Six Sigma and Shainin is, as discussed previously, that the overall DMAIC strategy implemented by Six Sigma is strong. The techniques used to implement the Measure and Analyse phases of the improvement cycle are weak, with a lot of the techniques relying on validating "guesswork" or using complex statistical tools. In contrast Shainin establishes a systematic approach to the analysis to eliminate variables in a process which do not have an effect on the overall variation. Rather than treating Six Sigma and Shainin as competing methodologies, aspects of Shainin can be blended with Six Sigma to produce a stronger and systematic approach which was proposed in the earlier section as the Hybrid Six Sigma method.

COMPONET A, B, C
and SUBSTRATE
ASSEMBLY

COMPONENT D and E
ASSEMBLY

OVEN and
SPRAY PROCESSES

MEASURING and
TESTING

Figure 34 Student Workstations in PIM game Manufacturing Process

The game simulates a manufacturing process with students acting as process operators. Assemblies are made by adding component parts to build up assemblies and then putting

them through a spraying and oven curing process, see Figure 34. The focus of the game is to demonstrate how the root cause of variation in a quality problem (known in Shainin terminology as the Red X^{SM}) can be identified and improved using the hybrid Six Sigma framework. The game starts with students defining a project which will return the greatest financial gain to company. Six Sigma performance metrics are then used to present the size of the quality problem in terms of defects, defects per unit, defects per opportunity, defects per million opportunities and Sigma level. Each assembly contains 4 critical-to-quality (CTQ) parameters known as faults; with each fault having a different root cause responsible for the variation seen in it.

The process problem is narrowed down by using the special data collection method known as the PVDT. The PVDT measures values defined as (CTQ) parameters on 4 consecutive samples at 5 separate times, with each sample measured 3 times. From these 60 results a Gage R&R study, a provisional Process Capability and a Shainin Multi-Vari study can be performed. This information will show whether the measurement systems are adequate, give us a baseline Capability index to assess the success of our quality improvement program and show the family of variation were the Red X^{SM} is present.

The game then progresses to demonstrate how to use this information to choose suitable tools, such as Isoplot^{SM}, BOB v WOW, Process Mapping and Cusum, for the follow up analysis to further narrow down the Red X^{SM}. Finally the game shows that by using a process of elimination, process variables which do not affect the product can be put to one side and reveals how to conduct a range of Design of Experiments like Variables Search, Components Search, Fractional Factorials and Full Factorials, with a reduced number of variables which allows us to conduct fewer experiments and can show up a greater number of higher level interaction effects between the important factors.

7.3   Developments

The current PIM game has evolved over much iteration, over many years. The original game concept was developed by John Cosier, a quality consultant, during the summer of 1998. Commissioned by Durham University after a one week course to teach MSc students Japanese problem solving tools was set up by John Wager in 1997. This course was successful but relied on a unique day of teaching from John Cosier on Shainin techniques, material he was exposed to whilst working for Philips in the early 1980's where he attended seminars led by Dorian Shainin.

The meeting to commission the game was conducted at Applied Hollographics in Newton Aycliffe. With representatives from Philips, ThyssenKrupp Tallent and John Wager, John Garside, Ernie Appleton and Mike Holgate of Durham University present.

The game was then used live for the first time in August 1999 by John Garside and John Cosier at Philips Durham. However the teaching material at this time did not incorporate the Six Sigma Methodology that was increasingly prevalent in industry. The need to include Six Sigma material became apparent when demonstrating the PIM game to Black & Decker Six Sigma black belts in 2001, where the Shainin material was well received but techniques such as Gage R&R would have to be included to fit with a Six Sigma organization.

In 2003 the PIM game was run including Six Sigma material developed by John Garside. This included techniques such as performance metrics, Gage R&R and Process Capability alongside the already established Shainin Material.

In Cox [29] modifications to the game were suggested in order to show students how to apply the demonstrated techniques to situations where there are both single-sided and double-sided specifications. As the game had previously only shown the situations were a single-sided specification was used.

7.3.1 *Testing of Modifications*

Part of this research was to test and refine these modifications to implementation. Developing the process improvement techniques demonstrated so their methods could be applied to the double-sided situation, as described in chapter 6. The modifications proposed in Cox [29] were trialled in a demonstration of the new PIM Game at the Industrial Statistics Research Unit, Newcastle University. This provided an opportunity to fully test the practical implication of the recommended improvements.

During this trial, out of the four faults that students test in the PIM game faults 2 and 4 ran without issue and provide a platform to demonstrate specific Shainin six sigma tools and techniques. In the case of fault 2 this was unsurprising, the mechanics of this fault remained the same as the previous tested iteration of the PIM game; providing a fault with a single sided specification, an adequate measurement system demonstrated using Gage R&R, clear Red X^{SM} which is a piece-to-piece problem highlighted with a Multi-Vari study and a BOB vs WOW investigation showing nothing clear. Finally using process mapping it is possible to uncover the best settings of pressure and viscosity for this fault.

Fault 4 however had new mechanics to provide a fault with a double sided specification; with a Gage R&R score which was inadequate and the associated within-piece Red X^{SM} visible in a Multi-Vari study. Using an Isoplot^{SM} the students demonstrated the large variation was a result of an inadequate measurement system. Following the implementation of an improved measurement system, full factorial experiments showed that underneath large measurement system variation there was an interaction effect between the substrate used in the assembly and component B. The responses described were as expected to provide a platform to demonstrate the specific techniques needed to highlight the root causes of variation in the fault.

Fault 1 demonstrated a few teething problems that need adjustment in order to fully demonstrate process improvement tools with the new double sided specification. During the start of the game two batches of assemblies are produced in order to produce six sigma metrics such as DPMO for the process. At this stage of the trial game fault 1 produced assemblies that all passed the quality check stage, which was undesirable and not intended. This would lead a normal six sigma investigation to determine the process as a good one. As the game is intended to demonstrate process improvement techniques, it is required that this fault produces enough failed assemblies so that the process is determined to have a sigma level significantly lower than a Six Sigma process. There should be failed assemblies produced either side of the process target so the students can determine the range of the variation seen. There were two bugs in the game which resulted in this problem; first component A has a large range between 4-10. This allowed the component A to have a larger effect on the fault 1 score than was desired. It was intended that this fault would have a large main effect created by component C and component A would have a smaller effect interacting with C. This range was allowing assemblies with components C that were expected to cause the assembly to fail, to pass. Secondly the initial values for component C that game produces for the first two batches of assemblies is fixed to produce the required balance of good and bad units. The range of variation is determined from the second of these initial batches and was intended to produce a bad assembly above the upper specification limit and a bad assembly below the lower specification limit; this did not happen as a result of the initial component A problem. This also raised the question if two fault 1 fails per batch of nine assemblies were enough.

During the second phase of the PIM game a Shainin Multi-Vari study is used to demonstrate a clear time-to-time Red X problem however this did not show up in the trial. This effect is created in the game by the C values changing at a set frequency from random values that will cause bad fault 1 assemblies above the upper specification limit, then bad

fault 1 assemblies below the lower specification limit and finally good fault 1 assemblies. The change happened every 12 components, the rationale behind using 12 and not using 9 or 18 which would correlate directly with one or two batches was to incorporate outlier results which are common in live data collection. However confounded with the previously described large effect of A this caused the time to time effect to be hidden underneath a larger piece to piece effect.

The undesirably large effect of A appeared again during the third phase of the PIM game. At this stage BOB v WOW is used to highlight that the best assemblies are produced when components C is in the range 9-11 of a possible range of 0-20. This did not occur in the trial of the PIM game; it was attributed as a direct consequence of the effect of A being too large and causing assemblies with C components outside of the range 9-11 to be good assemblies.

Fault 3 contained new mechanics in order to give its output a double sided specification. In the early stages of the game they appeared to work. Providing the correct number of pass and failed assemblies in the first two batches to provide the correct six sigma performance metrics. The Gage R&R also showed an inadequate score with this reflected by the within-piece Red X^{SM} variation in the Multi-Vari study. The Isoplot^{SM} then showed a large amount of product variation but little variation in the measurement system, therefore demonstrating that despite the poor Gage R&R score the measurement system was valid. An issue arose whilst using a BOB vs WOW technique on this fault, this should show up the best unit are produced when components E is set to 3 within a range of 1-5. This did not happen in the trial with a range of the component E values showing up as the best.

7.3.2 *Resolution of Issues from Testing*

To resolve these issues the formula for fault 1 proposed in Cox [29] shown in (6) was examined. The first section of this formula, highlighted in (7), is the part which controls the interaction between component A and C in fault 1 of the PIM.

=IF(C>12,110,IF(C<8,90,100))+((A-7)*2*IF(A<7,-1,1)*(C-10))+(10-D)+(2-RAND()*4)

$$(6)$$

=IF(C>12,110,IF(C<8,90,100))+((A-7)*2*IF(A<7,-1,1)*(C-10))   (7)

The distribution of the output from (7) is plotted within the confines of A in the range of 4-10 and C in the range of 1-3, 9-11 and 18-20 in Figure 35. The ranges for A and C were taken from the ranges used in the trial game. We can see that when A=7 the potential fault

1 score is within 10 units of the target value of 100. It is clear from this that once the effect of component D and the random variation effect present in (6) is added in, this can cause a further shift in the range -2 to 12 in the fault 1 output. This is enough to cause assemblies with component A = 7 and Component C = 1, 2 or 3 to have fault 1 scores equal to the process target of 100. This resulted in the confusion in the BOB v WOW test results, which should show clearly that all units close to the process target of 100 should have a component C = 9, 10 or 11.
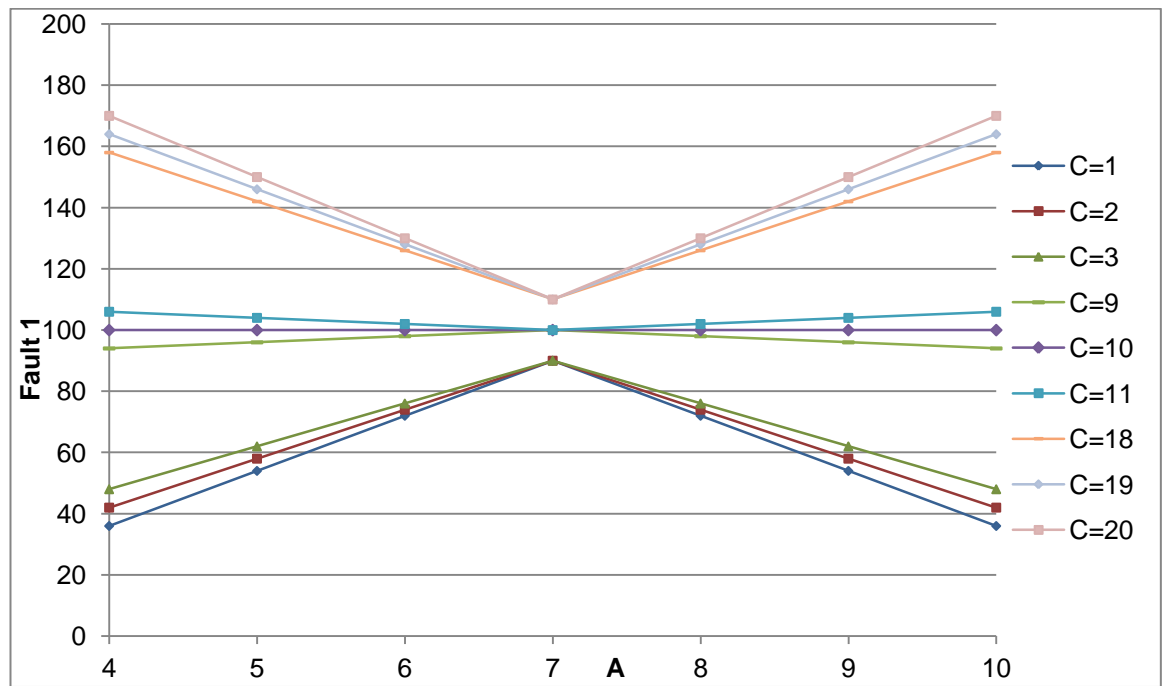
Figure 35 Distribution of Fault 1 results from (7)

From reviewing this information it was decided to keep fault 1 as a double sided specification but that component A would revert back to a single sided specification. Given that the PIM game is aimed at teaching process improvement methods in the classroom; over complexity of the component effects in a fault, which have a double sided specification appear to be too time-consuming in the trial to unpick. So the revised formula for fault 1 will reduce component A to a single sided specification and increase the minimum margin between good component C values of 9-11 and the bad component C values of 1-3 and 18-20.

$$= IF(C>12,75,IF(C<8,120,100))+((A-10)*1.25*(C-10))+(10-D)+(2-RAND()*3) \quad (8)$$

The revised formula for fault 1 is shown in (8) and the distribution of the new AC interaction is shown in Figure 36. The distribution of the AC interaction in Figure 36 demonstrate that there is now a clear gap of 20 units in the boundaries between when the values of component C are good and bad. This should give sufficient movement in the fault 1 output, even when the small effects of component D and the random variation are

factored in, stopping assemblies with component C values which are bad showing up as good in the BOB v WOW test. It is also recommended that the Component A input range is reduced from 4-10 to 6-10.
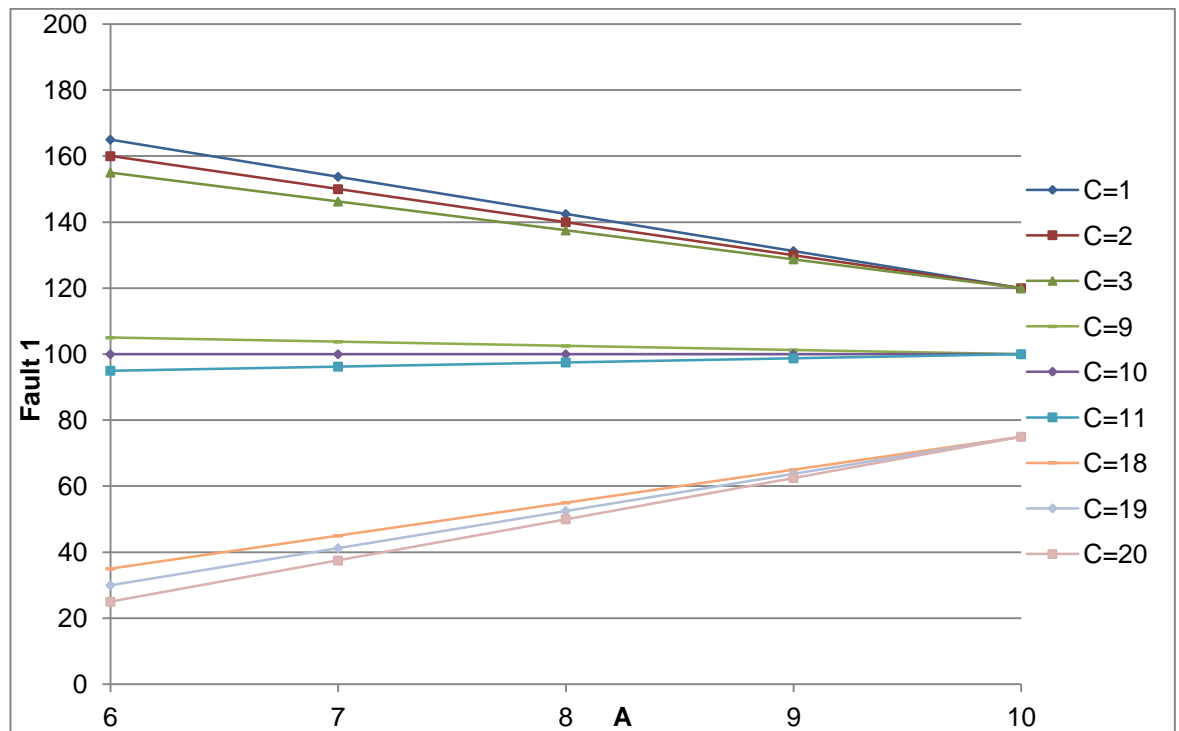
Figure 36 Distribution of Fault 1 results for A*C Interaction from (8)

Once the new formula was finalised, the issue of fixing the C value inputs firstly to give the required performance metrics at the start of the game and secondly to deliver a clear time-to-time Red X^{SM} in the Multi-Vari study.

In order to give the required performance metrics at the start of the game the first batch of nine assemblies must contain four failed assemblies and five passed assemblies. In the second batch there must be five pass assemblies, two failed assemblies at the lower end of the fault 1 distribution and two failed assemblies at the upper end of the fault 1 distribution. This is achieved by fixing the first 18 available values component C, such that the first five numbers available are in the range 9-11 (this will give five good assemblies in the first batch), the next six values should be in the range 18-20 (this will give four bad assemblies in the first batch and two bad assemblies at the upper end of the fault 1 distribution in the second batch), then two values in the range 1-3 (this will give two bad assemblies at the lower end of the fault 1 distribution in the second batch) and finally another five values in the range 9-11 (this will give five good assemblies in the second batch).

To resolve the Multi-Vari study problem the frequency with which the values of the component C change will dictate the clarity and family of problem shown up in the Multi-

Vari study. For the trial game the component C values changed every 12 parts, which meant that the game produces 12 parts in the range 9-11, 12 parts in the range 18-20 and 12 parts in the range 1-3. In the game batches are made in groups of nine, the rationale behind changing every 12 parts was to introduce some random outliers to the Multi-Vari study which is common in life situations. However frequency change of 12 parts was creating too many outliers to the extent that the Multi-Vari study demonstrated that the Red X^{SM}. variation was piece-to-piece rather than time-to-time. As five batches are tested during the Multi-Vari stage, it was decided to change the frequency to 18 parts; this should mean the first two batches are all good with component C values in the range 9-11, the following two batches will be bad in the upper range of the fault 1 distribution with component C values in the range 18-20 and the final batch will be bad in the lower range of the fault 1 distribution with component C values in the range 1-3.

To resolve the issue with fault 3 the formula that was used for fault 3 in the trial, showed in (9), was reviewed. Firstly the random variation effect was removed from the formula as in (10) and then the distribution of the oven and component E interaction was plotted in Figure 37.

$$=88+((20-\text{OVEN}_{corrected})*(3.5-E)*\text{IF}(\text{ABS}(3-E)=0,12,\text{IF}(\text{ABS}(3-E)=1,6,4)))+(3-\text{RAND}()*6) \tag{9}$$

$$=88+((20-\text{OVEN}_{corrected})*(3.5-E)*\text{IF}(\text{ABS}(3-E)=0,12,\text{IF}(\text{ABS}(3-E)=1,6,4))) \tag{10}$$

Figure 37 demonstrates were the potential issues with the BOB vs WOW test on fault 3. Showing if component E is at its best setting of 3 but interacting with oven positions such as 10, which is a poor oven setting, this will produce a fault 3 result which is further away from the process target of 100 than if component E is at a poor setting of 1 or 2 and the oven position is a good setting of 8. This gave the confusing Bob versus wow results in the trial game which was intended to highlight that component E is the most important factor.

There are two options in order to resolve this issue; the first method would be a slight adjustment of the proposed formula shown in (9). This adjustment can be seen in (11) and is plotted in Figure 38.
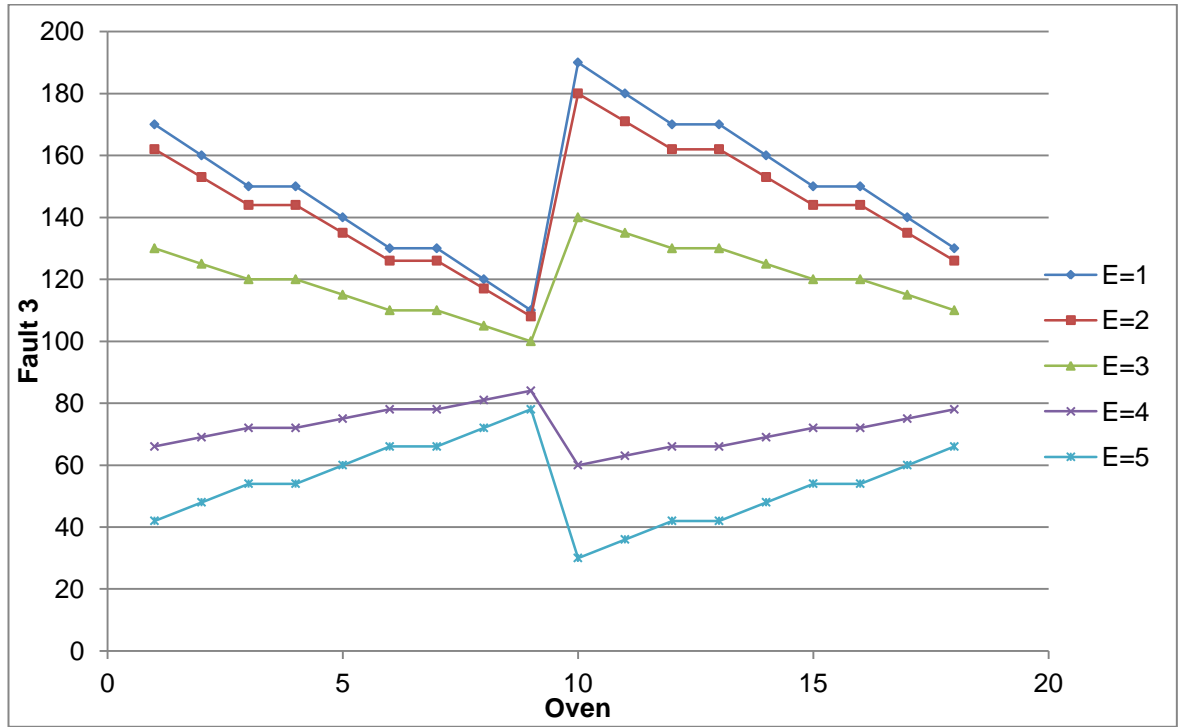
Figure 37 Distribution of Fault 3, Plotting Oven*E interaction from (10)

$$=88+((20-OVEN_{corrected})*(3.5-E)*IF(ABS(3-E)=0,8,IF(ABS(3-E)=1,10,4)))+(3-RAND()*6) \tag{11}$$

The adjustment made in (11) is to the middle part of the formula, IF(ABS(3-E)=0,8,IF(ABS(3-E)=1,10,4)), which acts as a scaling factor to change the size of the effect of E based on its value. By reducing the size of a scaling effect, from 12 to 8, when component E is at its best setting of 3 and increasing the scaling effect, from 6 to 10, when component E is equal to 1 or 2; this will move the fault 3 values closer to the process target of 100 when component E is set to 3. It will also move the fault 3 values further away from the process target when component E is set to 1 or 2. This change would make a BOB vs WOW a lot clearer and the reduced overlap between the fault 3 values at different settings of component E, as plotted in Figure 38.

The second method of resolving the issue with this fault would be to revert back to the previous tested iteration of the PIM game. In comparison to the first possible resolution this may initially be seen as a negative solution as it will not provide a fault with a double sided specification as was intended in Cox [29]. The reason behind the changes proposed by Cox [29] and tested at ISRU was to provide a game which contained double sided specifications to teach process improvement methods with, as the previous iteration only contained single sided specifications. Therefore converting all faults to a double sided specification will not improve the game, but merely demonstrate to students' application of techniques to the double sided case and not the single sided. Leaving a balance of single

64

sided and double sided specifications would improve the richness of the game. Hence if the mechanics of Fault 3 returned to the previous version, shown in (12), this will provide faults 1 and 4 with a double sided specification and faults 2 and 3 with single sided specifications. This will give the students equal opportunity to apply process improvement methods to both single and double sided specifications, which will be important as in industry it is common to find both types of specification.

$$=\text{Oven}_{\text{CORRECTED}}*E+10+RAND()*6 \tag{12}$$
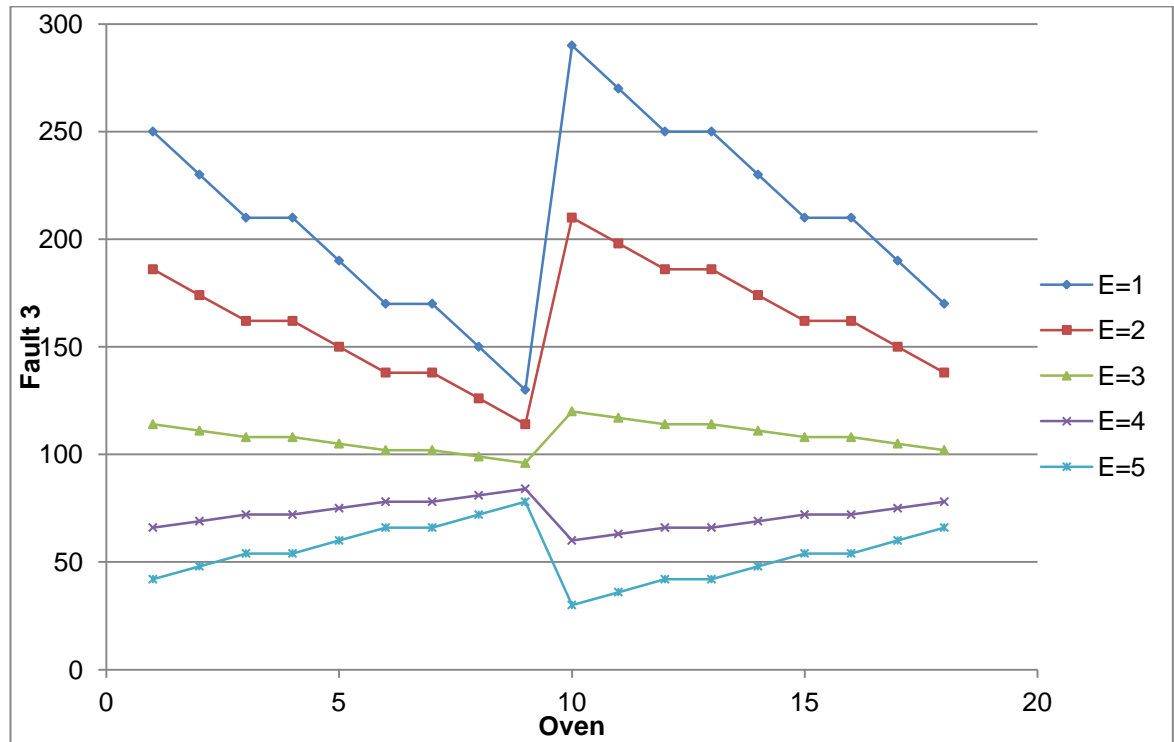
Figure 38 Distribution of Fault 3, Plotting Oven*E interaction from (11)

7.4 Factorial Experiments

A final point that was drawn out of the development of the two sided specification faults is that the students must take a careful approach to Factorial Experiments. When performing Components Search, Variable Search, Fractional or Full Factorial Experiments on a two sided specification the students must decide whether to analyse the results as produced or correct them and perform a distance from target approach. If the latter is taken the students must use real values and not absolute values i.e. if a result from a fault is 70 and the target value is 100 the result used in the Factorial should be -30 not 30.

Table 8 demonstrates a ANOVA table of result from a full factorial experiment on fault 1, using the results directly output from the measurement process. This shows that factor C has the most significant effect. Table 9 demonstrates an ANOVA table containing the same output from fault 1 as Table 8 except the results are adjusted using the target is best

method using real values. Once again demonstrating Factor C is the most significant effect.

However Table 10 demonstrates what happens to the factorial analysis of the fault 1 results if they are adjusted to target is best using absolute values. Factor A is incorrectly diagnosed as the most important effect, this has been caused by the difference between the average results for factor C when it is positive and negative been reduced from 740 to 16.

To ensure this mistake is not made by student in live projects, it is recommended that the PIM game demonstrator uses the correct target is best method highlighted in Table 9 on fault 1. Then demonstrate the direct results method highlighted in Table 8 on fault 4.

Table 8 Fault 1 Full Factorial using direct results

RUN	A	E	C	D	AE	AC	EC	AD	ED	CD	AEC	AED	ACD	ECD	AECD	RESULTS
12	-1	1	-1	-1	-1	1	-1	1	-1	1	1	1	-1	1	-1	175
16	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	173
15	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1	166
11	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	164
4	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	130
8	1	-1	-1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1	129
7	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	121
3	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	120
2	1	1	1	-1	1	1	1	-1	-1	-1	1	-1	-1	-1	-1	85
6	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	84
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	75
5	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	75
14	-1	-1	1	-1	1	-1	-1	1	1	-1	1	-1	1	1	-1	35
10	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	33
9	-1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	-1	26
13	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	25
																1616
LEVEL 1	819	808	438	772	809	997	808	807	806	808	809	807	807	812	805	
LEVEL -1	797	808	1178	844	807	619	808	809	810	808	807	809	809	804	811	
DIFFERENCE	22	0	-740	-72	2	378	0	-2	-4	0	2	-2	-2	8	-6	
EFFECT	3	0	-93	-9	0	47	0	0	-1	0	0	0	0	1	-1	

Table 9 Fault 1 Full Factorial using target is best real values

RUN	A	E	C	D	AE	AC	EC	AD	ED	CD	AEC	AED	ACD	ECD	AECD	RESULTS
12	-1	1	-1	-1	-1	1	-1	1	-1	1	1	1	-1	1	-1	75
16	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	73
15	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1	66
11	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	64
4	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	30
8	1	-1	-1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1	29
7	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	21
3	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	20
2	1	1	1	-1	1	1	1	-1	-1	-1	1	-1	-1	-1	-1	-15
6	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	-16
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-25
5	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	-25
14	-1	-1	1	-1	1	-1	-1	1	1	-1	1	-1	1	1	-1	-65
10	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	-67
9	-1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	-1	-74
13	-1	-1	1	1	1	-1	-1	-1	-1	-1	1	1	1	-1	1	-75
																16
LEVEL 1	19	8	-362	-28	9	197	8	7	6	8	9	7	7	12	5	
LEVEL -1	-3	8	378	44	7	-181	8	9	10	8	7	9	9	4	11	
DIFFERENCE	22	0	-740	-72	2	378	0	-2	-4	0	2	-2	-2	8	-6	
EFFECT	3	0	-93	-9	0	47	0	0	-1	0	0	0	0	1	-1	

68

Table 10 Fault 1 Full Factorial using target is best absolute values

RUN	A	E	C	D	AE	AC	EC	AD	ED	CD	AEC	AED	ACD	ECD	AECD	RESULTS
12	-1	1	-1	-1	-1	1	-1	1	-1	1	1	1	-1	1	-1	75
16	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	73
15	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1	66
11	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	64
4	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	30
8	1	-1	-1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1	29
7	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	21
3	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	20
2	1	1	1	-1	1	1	1	-1	-1	-1	1	-1	-1	-1	-1	15
6	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	16
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	25
5	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	25
14	-1	-1	1	-1	1	-1	-1	1	1	-1	1	-1	1	1	-1	65
10	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	67
9	-1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	-1	74
13	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	75
																740
LEVEL 1	181	370	362	370	369	359	370	371	366	406	369	373	371	372	371	
LEVEL -1	559	370	378	370	371	381	370	369	374	334	371	367	369	368	369	
DIFFERENCE	-378	0	-16	0	-2	-22	0	2	-8	72	-2	6	2	4	2	
EFFECT	-47	0	-2	0	0	-3	0	0	-1	9	0	1	0	1	0	

8   Destructive Test Measurement System Analysis

When Measurement System Analysis (MSA) is carried out in industry the Gage R&R is the most common test of the effectiveness of a measure [4]. The % Gage R&R of a measurement system is represented by the Xbar/R Method [16] using the formula in (13); where EV represents equipment variation, AV represents appraiser variation and TV represents the total variation of a measurement system. This test is a statistical assessment of how robust a measurement system is measuring variations in a product against the variation created by the equipment itself, or how repeatable the test is, and variation created by appraiser, or how reproducible the test is.

$$\frac{\sqrt{(EV^2 + AV^2)}}{TV} = \%GR\&R$$

(13)

For MSA to class a measurement system as acceptable it needs a % Gage R&R of less than or equal to 10% [16]. This is effectively stating that the total amount error in a measurement system should be less than 10% of the total variation recorded by it.

The effectiveness of this test is reliant on the assumption that the product being measured is robust against the test method itself. In earlier sections it has been described how an Isoplot[SM] test can be used to test this assumption, i.e. if the test method has an effect on the product itself. This type of Gage R&R study therefore relies on a test being repeatable. If a test method breaks the repeatability assumption it is classed as a destructive test. New assumptions have been to be formed in order to calculate a Gage R&R of a destructive test method.

De Mast [30] offers a summary of alternative assumptions and methods to determine a destructive Gage R&R score. The most common assumption to make when performing a destructive Gage R&R is to assume homogeneity between consecutive units sampled. This assumption is fine when there is little or no variation between consecutive units tested as this variation will be confounded with the repeatability or equipment variation. Therefore methods where homogeneity cannot be assumed will be focused on.

In this section two of the methods offered will be elaborated on and examples of their use described in the subsequent section. The first method that will be looked at would be to use a second measurement system which is impractical for regular use but can be used to validate the Gage R&R of the primary measurement system which is destructive. The second method will be to use a reference sample of known consistency to validate the destructive measurement system.

8.1 Alternative Measurement System

The basic premise of this approach is where there is a situation in which homogeneity between consecutive units cannot be assumed. Therefore doing so would lead to an overestimate of the Gage R&R as within variation between consecutive units is confounded with the repeatability of the measurement system. If there is a second non-destructive measurement system available for experimental purposes even if it is not practical for continued use, De Mast and Trip [30] describes how the non-destructive method can be used to assess the size of the overestimate. In the text De Mast and Trip applies the principle using the ANOVA method, in this section this method will be interpreted to use the Xbar/R method. Although Xbar/R method is considered by statisticians as less accurate than the ANOVA method, it is widely used in industry as results produced by this method are easier to interpret and understand.

The procedure that was developed starts by collecting sample units from production to be measured by the test equipments such that:

- Even numbers of consecutive units are collected (typically 4), as consecutive units will represent repeated measurements in the destructive test. They must also be enough units for more than one operator to test more than once each batch.
- Consecutive units collected at different sample times in order to give a total number of 60 parts collected, over a time period long enough to capture at least 80% of the variation seen historically.

Once the sample units have being collected, a standard non-destructive Gage R&R will be organised on the units as follows:

- Select one unit randomly from each batch of consecutive units. If consecutive units were collected in batches of four, this will give a total number of 15 units for the non-destructive Gage R&R.
- Each unit should then be measured by number of operators (typically 2), for a number of repeats (typically 2) so that 60 measurements have been taken.
- The results should then be processed using either statistical software, for example Minitab, or by hand using the Xbar/R Gage R&R method described in the QS9000 Measurement System Analysis reference manual [16]. The final calculation is shown in (14), with the subscript ND to represent the non-destructive test.

$$\frac{\sqrt{(EV_{ND}^2 + AV_{ND}^2)}}{TV_{ND}} = \%GR\&R_{ND}$$

(14)

The next stage of the procedure is to measure all other collected units using the non-destructive test, in the example of 60 units this will mean 60 measurements. The

measurements should be made by two operators such that each operator is measuring half of the units collected in each batch of consecutive units. The order in which the samples are tested should be randomised and also recorded so that the units can be tested destructively between two operators in the same order.

This new non-destructive Gage R&R effectively assumes homogeneity between parts in the same manner a destructive test would. Formula (15) demonstrates the Gage R&R arrangement, with WIV representing the within variation caused by the variation between consecutive parts and DR&R representing the Gage R&R score for the destructive test arrangement containing the within variation.

$$\frac{\sqrt{(EV_{ND}^2 + AV_{ND}^2 + WIV_{ND}^2)}}{TV_{ND}} = \sqrt{\frac{EV_{ND}^2 + AV_{ND}^2}{TV_{ND}^2} + \frac{WIV_{ND}^2}{TV_{ND}^2}} = \%DR\&R_{ND}$$

(15)

Once this is calculated the size of the within variation can be estimated as in (16) by subtracting (14) from (15).

$$\frac{WIV_{ND}^2}{TV_{ND}^2} = \%DR\&R_{ND}^2 - \%GR\&R_{ND}^2$$

(16)

The final stage of testing involves using the Destructive test method in the same arrangement as the previous stage of testing. Whereby the given example of 2 operators test 2 units from each batch of 4 consecutive units, the DR&R can then be calculated as in (17) with subscript DT representing the destructive test results.

$$\frac{\sqrt{(EV_{DT}^2 + AV_{DT}^2 + WIV_{DT}^2)}}{TV_{DT}} = \%DR\&R_{DT}$$

(17)

If (18) is assumed that the ratio of the within variation seen in the destructive test is approximately the same as the non-destructive test. The actual Gage R&R score for the destructive method can be approximated by substituting (16) from (17) as in (19).

$$\frac{WIV_{ND}^2}{TV_{ND}^2} \approx \frac{WIV_{DT}^2}{TV_{DT}^2}$$

(18)

It is worth noting that the strength of this approximation can be checked by a regression analysis of the destructive and non-destructive test results to check they are correlated.

$$\%DR\&R_{DT} = \sqrt{\frac{EV_{DT}^2 + AV_{DT}^2}{TV_{DT}^2} + (\%DR\&R_{ND}^2 - \%GR\&R_{ND}^2)}$$

$$\%GR\&R_{DT} = \sqrt{\%DR\&R_{DT}^2 - (\%DR\&R_{ND}^2 - \%GR\&R_{ND}^2)}$$

(19)

8.2   Reference Sample Test

Using reference samples of a known characteristic in place of a test sample is a method recognised by De Mast and Trip (2005) as being able to address the repeatability issue associated with destructive test validation. The key to this method is producing reference pieces which are consistent and are measure by the measurement system in the same range as the products that are normally measured.

9 TRW Destructive Test MSA

The scope of this project was follow on from the previous TRW project were a non-destructive peel test of FET weld strength was proven to be destructive. Therefore a second experimental but destructive sheer test was to be developed and validated. Although this test suffers from material waste it offered a very quick method of determining the capability of the laser weld process when compared to the microscope method.

This project provided the practical challenge of developing a finalised test rig design, as well as the statistical challenge of validating the repeatability and reproducibility of a destructive test method. Where destructive tests do not hold the assumption of the unit measured being robust against the measurement process as is normally assumed in a classic Gage R&R.

In order to validate this test equipment, two methods were pursued in order to determine a Gage R&R score. Due to known variation in the FET laser weld process the normal assumption of homogeneity cannot be made, so a reference sample will be attempted to be made and in parallel a non-destructive measurement of weld area on the FET will be used to estimate the size of the within variation before performing a destructive Gage R&R.

9.1 Alternative Measurement System

During preliminary investigations it had been uncovered that there may be a link between the maximum load results reported by the experimental sheer test (Figure 40) and the weld area on a FET laser weld measured under a Nikon Light Vision camera. To measure the weld area on each FET laser weld the height and length of both weld seams were measured and then the weld area calculated as a product of length and height, see Figure 39.
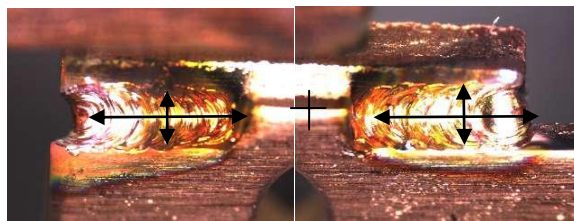

Figure 39 Photograph of a FET laser weld, marking the weld area

This link would first be tested to ensure that the experimental sheer test was practical to run on the factory floor at TRW and the slow to perform light vision camera measurements correlated with these result. If they do the assumption that the ratio of within variation will be the same in both tests can be made and followed with a more rigorous experiment.
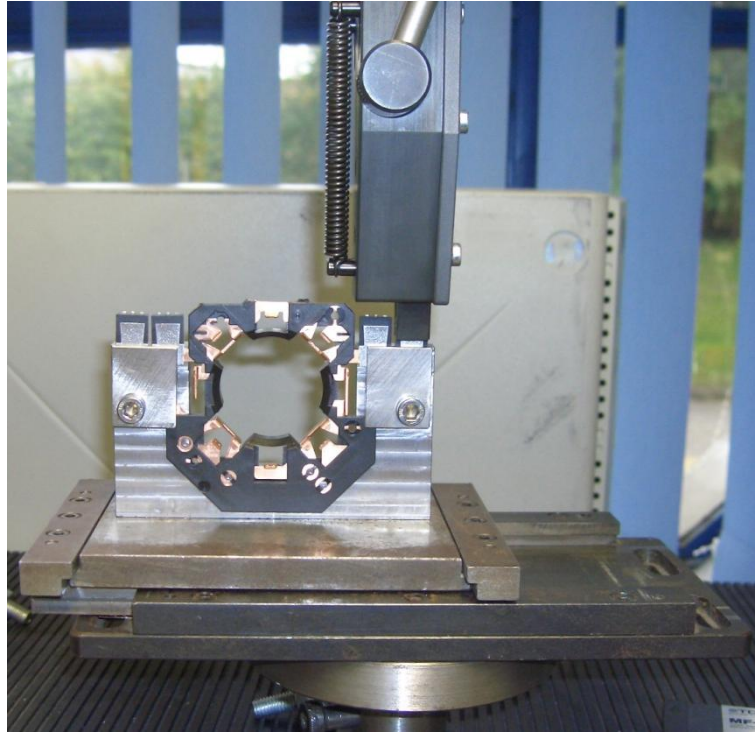
Figure 40 Photograph of a FET laser weld being tested in the Experimental Test Jig by a materials test machine

9.1.1 Factory Floor Gage R&R

This study was conducted on the factory floor to simulate working conditions, using the Testometric materials test machine. In this trial Homogeneity was assumed to provide repeat measures, but this assumption was subject to test. This was achieved by using the Nikon Light Vision equipment available in the TRW test lab. Using this equipment allowed all laser welds to be photographed, as in Figure 39, to allow a visual inspection of homogeneity as well as measurements of the total weld area for each FET.

This Gage R&R followed the structure of collecting samples of 4 consecutive units from the same stamping grid production cavity, Laser Welder and Laser Welder Nest. This trial used 16 samples to give 64 measurements, as 60 measurements are normally required to give confidence in a Gage R&R result. The 16 samples were collected from the following locations:

- 4 consecutive samples from Laser 1 nest 1, giving the largest strengths historically
- 4 consecutive samples from Laser 2 nest 1, giving a mid-range strength historically
- 4 consecutive samples from Laser 2 nest 2, giving a mid-range strength historically
- 4 consecutive samples from Laser 3 nest 1, giving the lowest strengths historically

All welds were photographed for visual inspection and weld areas measured using the Nikon Light Vision camera, to ensure consistency between FETs on consecutive samples. Then all destructive tests were carried out on the factory floor using the Testometric materials tester and experimental sheer test jig as in Figure 40.

75

The samples were randomised for testing by 2 TRW operators. Consecutive parts were classed as repeated measures and separate FETs classed as individual parts.

Comparing groups of 4 consecutive FETs revealed that visually these FETs shared similar characteristics in terms of the number of weld spots and shape of the weld seam. However when measuring the total weld area on the A and B weld seams on each FET combined, there appeared to be some variation. The variation between consecutive parts was smaller than the variation between the parts.

The results from destructively testing the same samples show the variation is greater part-to-part than between consecutive samples this is consistent with the Weld Area measurements. This seeming similarity between the results is confirmed in the regression plot in Figure 41 which has a Pearson correlation coefficient of 0.846, suggesting a strong positive correlation.
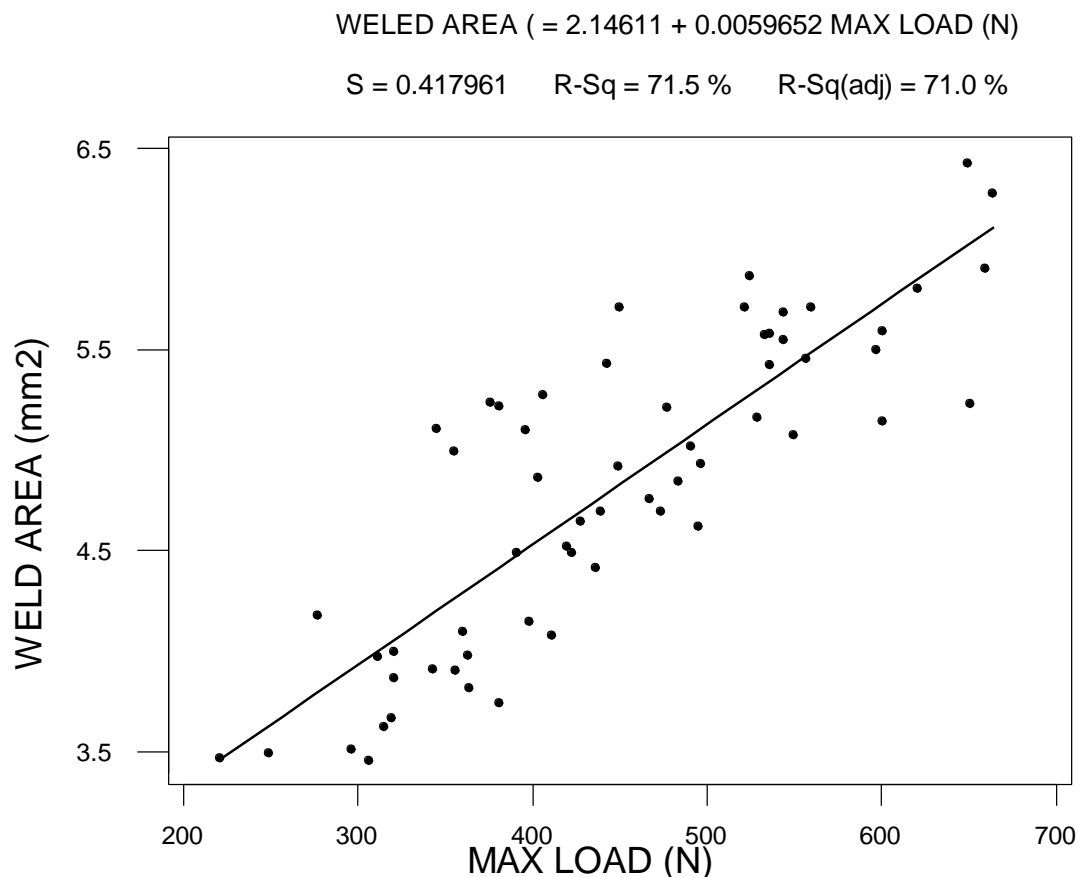
WELED AREA (= 2.14611 + 0.0059652 MAX LOAD (N)

S = 0.417961   R-Sq = 71.5 %   R-Sq(adj) = 71.0 %

Figure 41 Regression plot of Weld Area (mm2) against Max Load (N)

The variation between consecutive parts does have a detrimental effect on the statistics, where the Gage R&R = 36.63%. This would normally categorise the measurement system as inadequate. However Figure 42 clearly shows there is little variation attributed to reproducibility. Given product variation is confounded with repeatability variation and

therefore hidden in this Gage R&R statistic. Although at this stage it was difficult to assess the size of this product variation it is fair to say this value is clearly an underestimate.
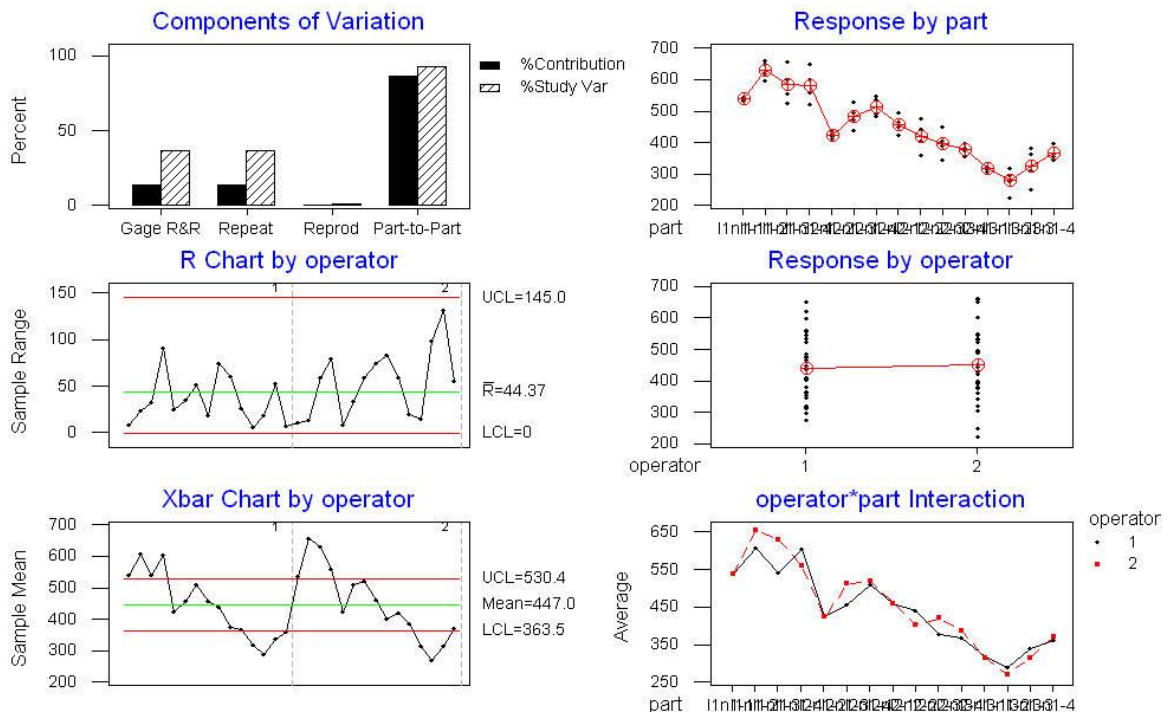

Figure 42 Destructive Gage R&R for the experimental sheer test

From this trial it was clear there was a link between the non-destructive Light Vision Camera measurements and the experimental destructive sheer test. From this a second trial was run this time determining the Gage R&R statistic for the Light Vision Camera to estimate the ratio of within product variation.

9.1.2 *Correlation between Light Vision and Destructive Test*

In order to estimate the size of the within-product variation confounded in the %Gage R&R statistic. An experiment was designed so that from one set of samples the actual %Gage R&R of the Light Vision Equipment can be assessed. Calculating the size of the within product variation by measuring the %Gage R&R for the Light Vision as if it were a destructive test and finding the difference between the squares of these results, as described in the previous section. This difference can then be removed from the destructive test result to determine the actual %Gage R&R score.

For this experiment all samples were collected from same cavity and all from Laser 3 Nest 1, to ensure consistent welds, as more consistent welds are easier to measure under the Light Vision Camera. Batches of 4 consecutive stamping grids were sampled, at the

following lamp age compensation power settings to simulate normal production variation; 88%, 90%, 92%, 94%, 98% and 100%. I.e. one laser welder was used to produce all the parts but the power of the welder was varied for each batch to simulate the variation of weld strengths in production.

The first stage involved collecting one sample from each power setting batch of four. This gave 6 samples to test, each with four laser welds and to be tested twice by two operators. The resulting 96 measurements under the Light Vision Camera were used to calculate a Gage R&R statistic of $\%GR\&R_{LV} = 22.44\%$. Where the subscript LV denotes the Light Vision Camera. The type of variation seen is plotted in Figure 43.
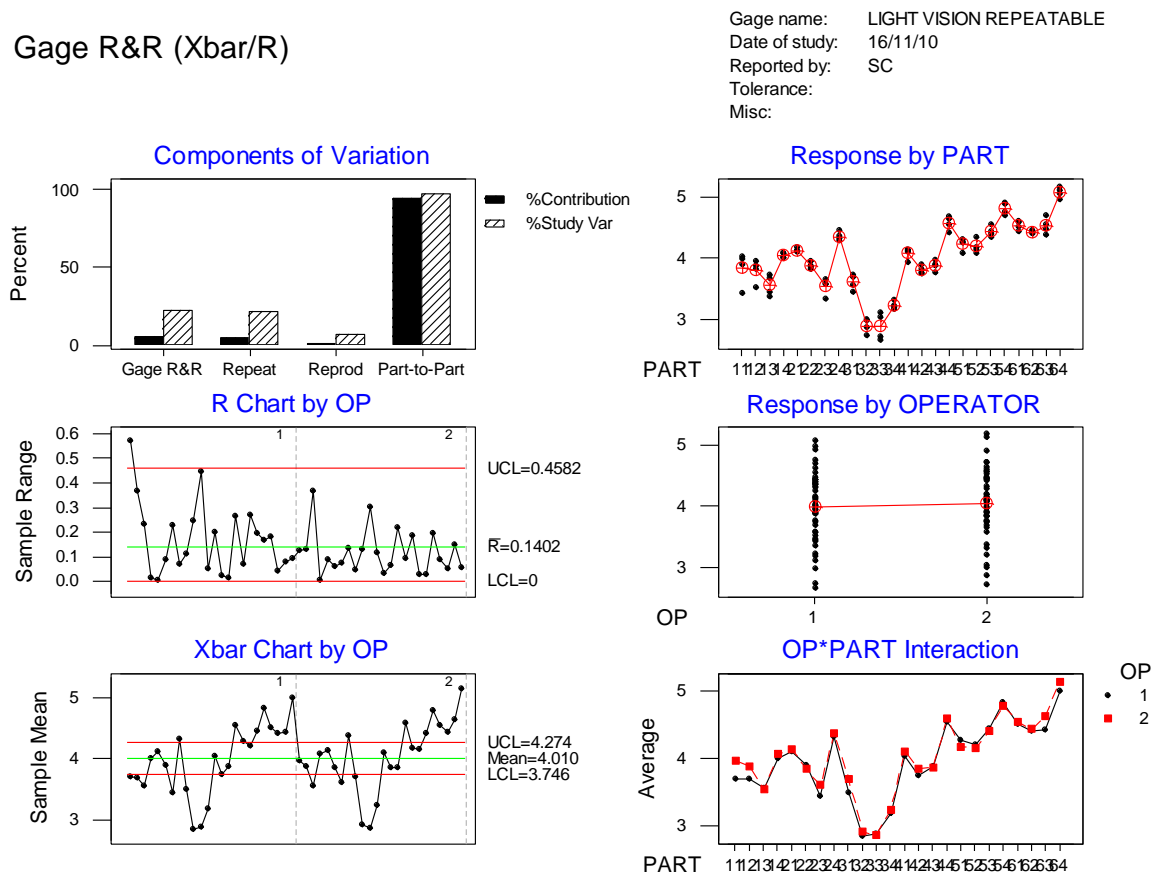

Figure 43 Non-Destructive Gage R&R using light vision Camera measurements of FET laser weld areas

In the second stage all the collected samples are tested non-destructively using the Light Vision Camera, each batch of four consecutive samples are split between two operators randomly to give a destructive test arrangement. This gave a destructive Gage R&R statistic for the Light Vision Camera of $\%DR\&R_{LV} = 40.56\%$.
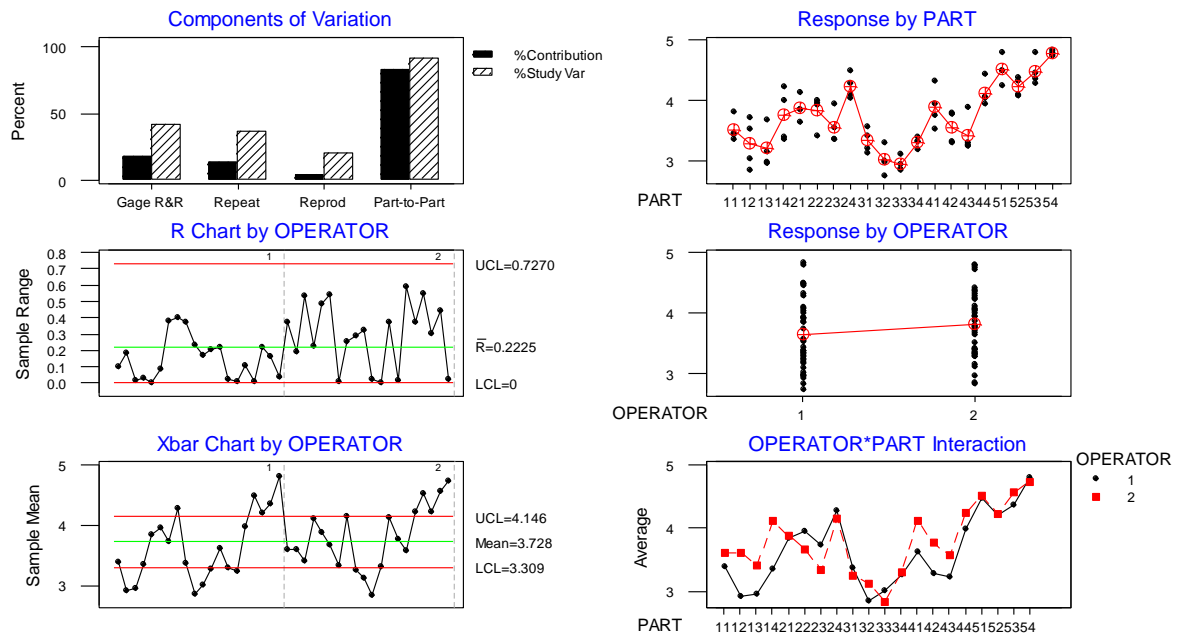
Figure 44 Destructive Arrangement Gage R&R using Light Vision Camera measurements of FET laser weld areas

The third stage involves destructively testing all samples with the experimental sheer test. The samples were tested between two operators in the same manner as the previous experiment. This gave the destructive gage R&R statistic for the experimental sheer test of %DR&R$_{ST}$ =38.44%. Where the subscript ST denotes the sheer test.

Figure 45 Destructive Gage R&R using Experimental Sheer Test of all samples

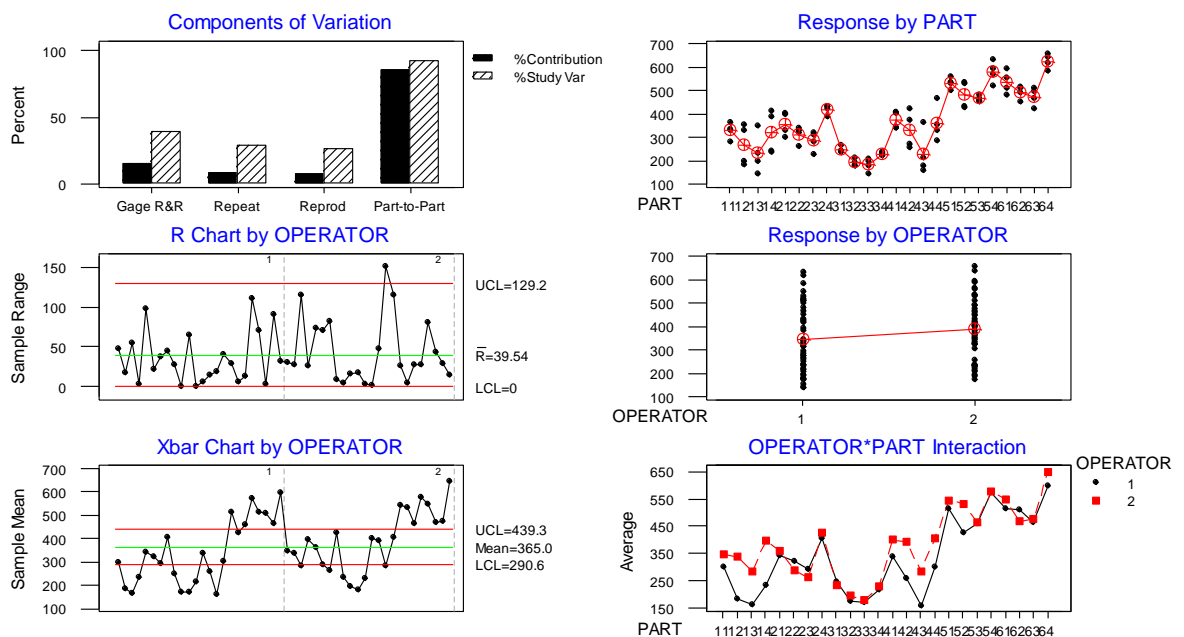The two sets of results from a destructive test arrangement were then plotted to check there is a strong correlation between the results in order to determine if the assumption that the ratio of within product variation is the same between both test methods. This showed there was a strong correlation with an $R^2=88.1\%$.

Regression Plot

MAX LOAD = -432.581 + 206.290 WELD AREA

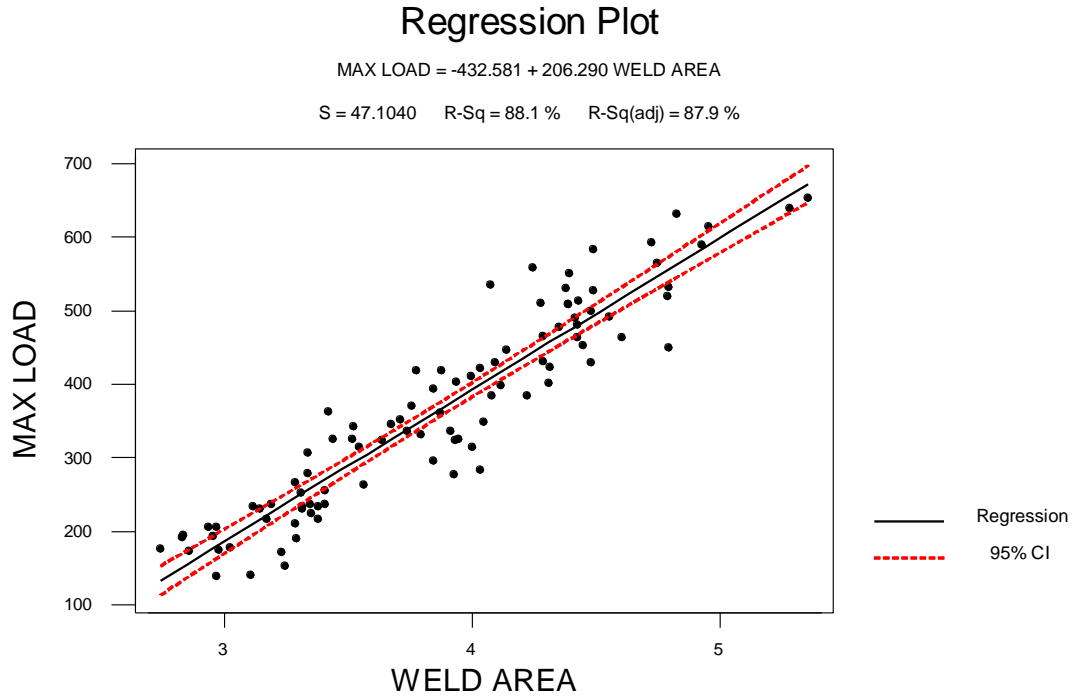S = 47.1040   R-Sq = 88.1 %   R-Sq(adj) = 87.9 %

Figure 46 Regression Analysis of the correlation between the maximum load recorded in the sheer test and weld area measurements recorded by the light vision camera

Using the method described in previous section estimate of the actual Gage R&R for the experimental sheer test was formed in (20). This result is interesting as it now classifies experimental sheer test as a marginal measurement system rather than inadequate measurement system.

$$\%GR\&R_{ST} = \sqrt{\%DR\&R_{ST}^2 - (\%DR\&R_{LV}^2 - \%GR\&R_{LV}^2)}$$

$$= \sqrt{38.44^2 - (40.56^2 - 22.44^2)} = 18.11\%$$

(20)

Interestingly this result appears to be skewed as the %AV scores for both $\%DR\&R_{LV}$ and $\%DR\&R_{PT}$ were around 20%. If this was removed the $\%GR\&R_{PT}$ would be reduced to 10.1% bringing it very close to the ideal. As this AV or reproducibility score has not been seen in previous trials and is of the same order in the $\%DR\&R_{LV}$ and $\%DR\&R_{PT}$ tests, it is suspected that this result has been caused by chance. It is thought the randomisation of

samples for testing on this occasion consistently resulted in one operator testing stronger samples than the other, rather than one operator testing in a different manner to another.

9.2 Reference Sample Testing

The possibility of using a reference sample to validate the test method was explored. Designing and making a reference sample provided a challenging technical problem. The samples would have to apply loads to the experimental sheer test jig in the same plain as the FET arms would when a Stamping grid is in the jig. They will also need to be produced in a material which is more consistent than the FET laser weld strength.

9.2.1 Spring

Using a specially made test piece (Figure 47) which hooks onto the experimental test jig in the same manner as the FET arms of a stamping grid token, a spring can be attached between the test piece and the top clamp. Repeatable measurements of extension of the spring can then be taken for varying loads which are representative of the range of maximum loads experienced when testing production pieces. This method did carry a big health and safety concern which lead to it being side lined. It was feared that if there was a failure of the jig, spring or test piece, the energy stored in the spring could lead to the potential of serious damage to the test equipment. Even worse would be serious injury to those standing near the test equipment if the energy in the spring was suddenly released.

Figure 47 Test Piece and Spring

9.2.2 Soldered joints

Soldering the FETs to the FET arms was a potential method of joining that is different to welding and may offer an improved consistency in bond strength. Figure 48 is an example stamping grid token with FETs soldered on.
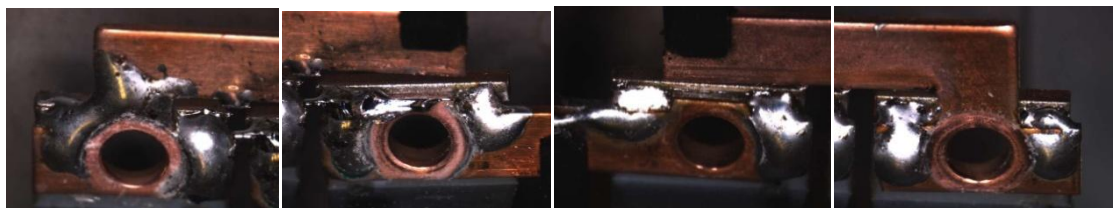

Figure 48 Solder Joints on FETs

Initial experimentation of this method was not encouraging. There was a big range in strength between 290N-1000N, which was counter to the objective of improved

consistency of the joint strength. Also many of the joints had over spill which needed to be filed away to fit the token in the jig. This filing process is inherently fraught as by definition changes the characteristics of the joint in an inconsistent way.

9.2.3   Machined Reference sample

A reference sample was designed that would theoretically break at a consistent load under similar conditions. Figure 49 shows the design of the reference sample which has been made to hook onto the lip of the jig simulating as closely as possible the forces exerted on the jig by the FET arms during the normal test conditions.
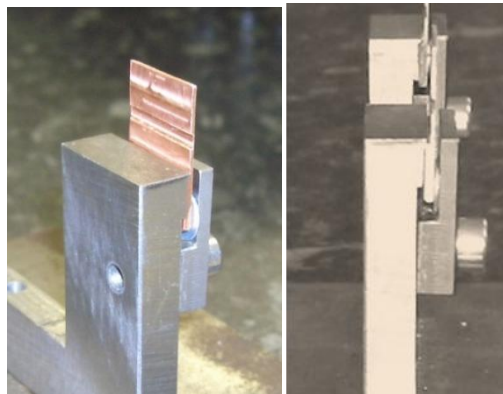

Figure 49 Photograph of machined reference samples

The test run results for the reference sample method demonstrated a surprisingly large amount of variation between test samples. Investigations of the samples lead to the discovery that there were random scratches on the surface of the break zone of the samples. When the samples were tested they were failing along the scratches as in Figure 50. Using the machining equipment available at TRW it is not possible to make consistent reference samples without the surface scratches. If this method was to be pursued further the production of the samples would need to be outsourced to a specialist company.
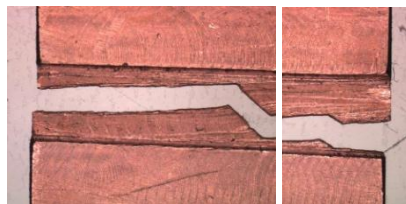

Figure 50 Photograph of break in reference sample

9.2.4   Copper Wire/Test Piece Hybrid

A method of testing a reference sample of cooper wire was tried. It was suspected that copper wire would be consistent between samples when compared with the machined reference samples. In order to apply forces to the jig through the copper wire in a manner

that simulate the FET arm resting against the lip of the jig, the test piece developed for use with a spring was used with a length of copper wire replacing the spring. (see Figure 51).
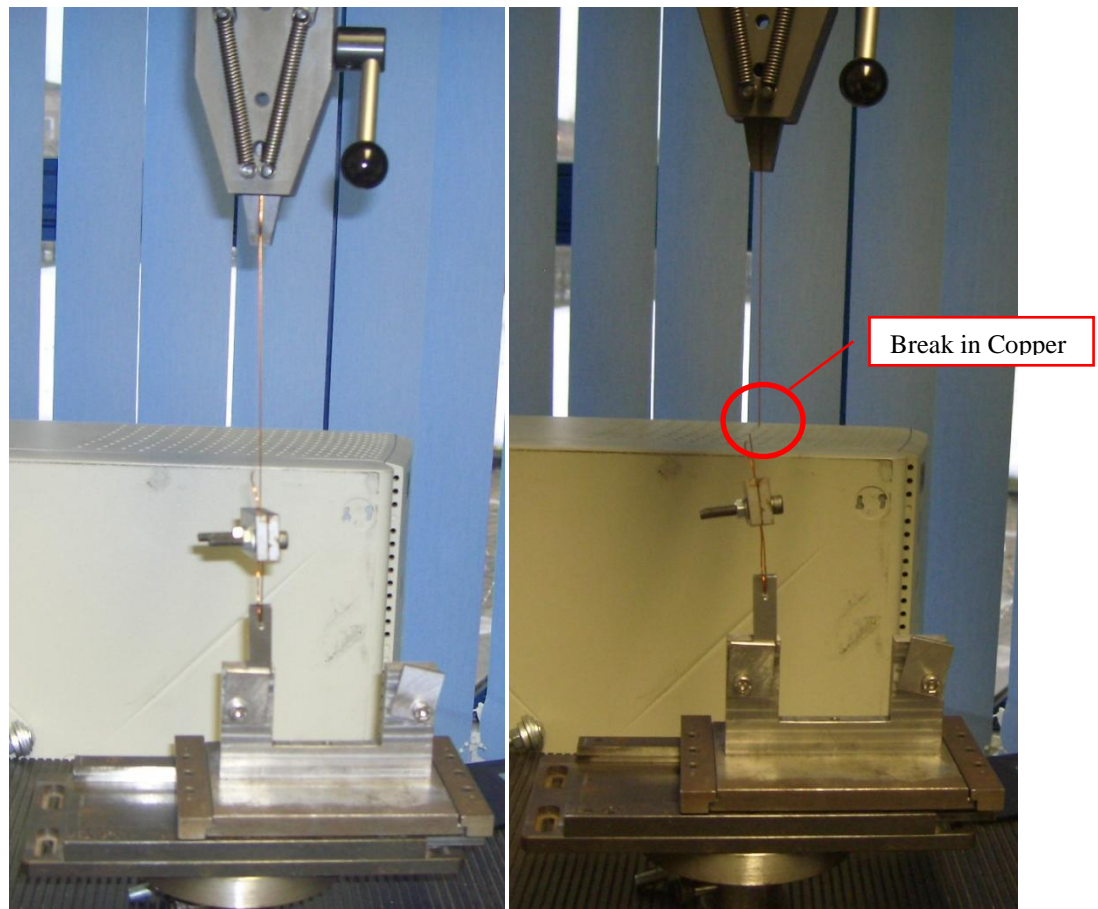
Figure 51 Photographs of Copper Wire test set-up

An initial trial was conducted using 0.670mm wire which was collected from scrap. The trial tested 30 pieces of wire giving a mean strength of 82N with a standard deviation of 2.61N.

Given the reasonably small amount of variation seen in the initial trial, wire of a range of thicknesses was ordered in to expand the test. Using Essex, TRW wire supplier, samples of wire that were 1.000 mm, 0.900 mm, 0.800 mm and 0.670 mm were ordered and used in a trial Gage R&R.

The Gage R&R was organised such that the left side of the jig was considered as an operator and the right side of the jig was considered as a second operator in order to compare any differences in the jig. Each thickness of wire was tested eight times in each side of the jig. This allowed for a total of 64 measurements in the Gage R&R study.

This reference sample test method produced a Gage R&R of 5.40%. This would categorise the experimental pull test method as adequate, as it is usual to consider any measurement system with a score of <10% as adequate. The results of the study in Figure 52 clearly

demonstrate there is little difference between test results obtained in the left hand side of the jig compared to the right. This gave us confidence that the jig will support the FET arms of FETs 1 and 2 in the same manner as the FET arms for FETs 3 and 4.

The only weakness in the method is that this selection of wires obtained only tested the jig in the range of 100N - 200N, whereas the strengths of the FET Laser Welds are usually in the range of 200N – 600N. Unfortunately there was not sufficient time to reorder thicker wire to repeat the test applying loads between 200N-600N, however if this was done and we saw the same variation of results for each wire over a bigger range this should improve the gage R&R score, unless there is a significant change in the jig performance at these higher loads.
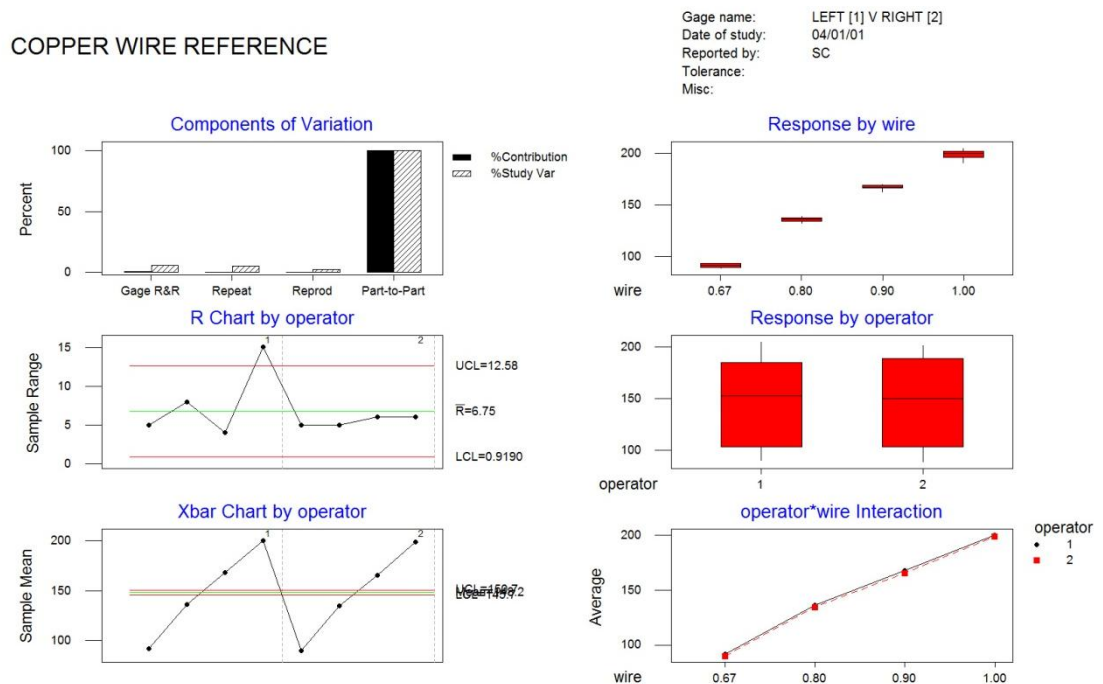
Figure 52 Gage R&R study for copper wire test

9.3 Discussion

Using stamping grids processed though the laser welders on the factory floor it has not been possible assuming homogeneity between consecutive parts, to determine the experimental pull test as an adequate measurement system. It is suspected that this is due to known process variation being confounded with the repeatability component of the Gage R&R statistic. However it was possible by referencing a second measurement system to determine an estimated Gage R&R which deemed the test as a marginal/adequate.

A gage R&R of 5.40% was finally determined for the experimental pull test after the development of a reference sample test which did not suffer from the confounding of

product variation with that of the repeatability of the test. This statistic validates the method as adequate, in accordance with QS 9000 standard.

The reference sample test only tested the Jig in the range 90N-200N however in operation the loads are likely to be in the range of 200N-600N. Due to time constraints it was not possible to order more copper wire to test the performance under higher loads. If the performance was maintained under these higher loads it will improve the Gage R&R statistic by increasing the total variation seen. From overall testing there did not appear to be any degradation of the jig which gives no reason that its performance will change at higher loads, as it has been exposed to these loads when testing with stamping grids.

Given the findings it was recommended that the experimental test method is implemented as it will improve confidence in the results obtained from it and allow for gains to be made tackling variation in the process.

10   Conclusions

10.1   Hybrid Six Sigma Approach

This thesis started by introducing two methodologies for process improvement; Six Sigma and the Shainin System. Similarities and differences, strengths and weaknesses were drawn between these two methodologies to derive a strengthened methodology. This provided context for the techniques and tools developed during the thesis.

The introduction highlighted two papers, Aboelmaged [1] and Senapti [2], which bring out the opportunity for academic research within the Six Sigma and Shainin System methodologies. Braking down then comparing and contrasting these methodologies lead to the development of a strengthened methodology called the Hybrid Six Sigma framework.

The Hybrid Six Sigma framework was formed from the realisation that there was little objective difference between Six Sigma and SS. This framework is based around the traditional 12-step Six Sigma DMAIC cycle given its strong overall structure and prevalence in industry. However a Shainin loop is included in the identification of variation sources of the Analyse phase. This added loop introduces the Shainin philosophy of "*finding the signature*" then "*narrowing down*" the potential root causes of variation. It also acts as conduit to introduces Shainin Techniques such as the Multi-Vari Study or Components Search into the mix of traditional Six Sigma Analyse techniques to bridge the gap between classically subjective and objective but complex Six Sigma tools.

10.2   PVDT; Efficient Diagnosis of the Source of Variation

The Hybrid Six Sigma framework developed was used as a basis for a sampling strategy known as the Process Variation Diagnostic Tool (PVDT). The PVDT provided a method whereby from 20 samples a Gage R&R and a Provisional Process Capability study can be found. It was structured to also provide an Isoplot^{SM} and a Shainin Multi-Vari study. Thus the obligatory DMAIC cycle steps of the Measure and Analyse phase are fulfilled, with a Gage R&R study and a Process Capability study. Whilst providing an early objective insight into the root causes of variation with the application of Shainin tools.

The method was then reviewed in three different industrial situations to demonstrate its effectiveness. The three cases represented different industrial applications; the first process produced panels for furniture from raw materials, the second assembled electrical components into electric motor control systems and the final low volume process produces microchannel plates to be later assembled in image intensifier night sights. All processes suffered from a chronic quality problem with one process losing £800,000 a year in lost production and another only achieving a 25% yield from the process.

Applying the PVDT to these problems allowed the project teams involved to quickly produce Gage R&R studies and Provisional Process Capability Studies. These metrics were required as part of the companies involved Six Sigma initiatives. However rather than performing two individual experiments these were obtained from one group of data produced by one experiment conducted within the PVDT guidelines. This dramatically increased efficiency, as it is common in a classic Gage R&R to take 60 measurements from typically 10 products and in a Process Capability Study to take 50 measurements from 50 products. Using the PVDT this was reduced from the combined 110 measurements from 60 products to 60 measurements from 20 products.

A significant advantage in the projects was the ability to extract a Shainin Multi-Vari Study from the 60 measurements taken for the PVDT study. This technique for the companies involved with the case studies was new to them but its benefits of improving objectivity in the early analysis of a problem became apparent. This technique allowed the project team, at no expense in terms of collecting new data, the ability to categorise the most significant families of variation. Finding this signature was strikingly beneficial in the first two case studies; the project team at the furniture manufacturer determined their panel process had a Red X[SM] within-piece problem. The company had previously suspected different size panels and changing process settings between batches was root cause for the variation seen; this hypothesis was quickly put to one side by the project team. If the edging and trimming machines were affected by the size of the panels or the adjustment of process settings there would be a significant batch-to-batch change in variation rather than the within-piece variation seen. This allowed the company to focus its efforts on uncovering the cause of within-piece variation; either a poor measurement system or the edging is uneven. In the electronic motor control systems manufacturer case study the project team uncovered a Pink X[SM] time-to-time problem. Following this up using a Cusum technique, the project team were able to link the changes in variation seen from time-to-time with changes in supply voltage from the Grid to the test equipment. At around 22:00 daily there was a spike in the supply voltage which was seen as an increase in the test results at the same time. This problem was resolved by installing voltage regulators between the Grid supply and the test equipment.

The second case study also highlights that the classification of families of variation is important as it aids the selection of follow up techniques. The ability to focus on a time frame of variation eliminates unimportant factors from the search for root causes. Knowing the number and type of suspect factors allows an objective view as to the type of statistical test or experiment can be used to further home in on the root causes of process

variation. The previous example showed how a time-to-time problem was successfully followed up with a Cusum graph. The third case study of a low volume process producing microchannel plates for image intensifier night sights demonstrates how a process with a large within-piece variation was followed up with an Isoplot^{SM} investigation. The Isoplot^{SM} tested the difference between Test Equipment 1 which is the current test rig and test equipment 2 which is an experimental test rig. The results showed two "sausages" of points either side of the 45° line on the Isoplot^{SM} graph which are specific to whether the channel plate was tested on equipment 1 first or equipment 2 first. Thus this plot shows that irrespective of which equipment is used the 2nd test always displays fewer SOC non-conformities. The Red X^{SM} within-piece problem was resolved by introducing an ageing process before the channel plates are tested, which lead to a significant increase in yield from 25% to 85%. It also meant output from the process increased without investment in new resources and the process became profitable. This also suggests that the majority of previously disposed channel plates with SOC problems would have passed inspection if they had been retested.

From these case studies it can be seen that at the border of the Measure/Analyse phase in Six Sigma the proposed PVDT offers an efficient method of collecting Six Sigma metrics and steering the course of an improvement project. In the second case study the in-house Quality Engineer commented that the project team had *"driven the project further in 2 weeks than it had been in the previous 2 years"*. Giving a clear indication as to the benefits using the PVDT had brought to the companies continuous improvement initiative.

10.3 TRW Case Study

An industrial project was conducted by the author in conjunction with TRW an electronic power steering systems manufacturer. The PVDT proposed in this thesis was used to validate an experimental non-destructive laser weld peel test procedure. This project differed from the previous case studies as they used the Isoplot^{SM} as a follow up to the PVDT. This thesis proposed a test method for the PVDT where an Isoplot^{SM} could be extracted from the results. The project highlighted how including an Isoplot^{SM} can speed up the diagnosis of weather a measurement system is valid or not and still maintain insights into the process through a Provisional Process Capability Study and Multi-Vari Study.

As parts where collected in trays of 12 parts on the factory floor, 24 sample parts where used for the PVDT rather than the proposed 20. Using these samples it was deduced with an Isoplot^{SM} that the measurement system was inadequate as a method of testing non destructively weld strengths of the FET laser welds. It was found that there were two

groups of results on the Isoplot^{SM}, which was demonstrating that the test was affecting the characteristic of the weld in a negative way. The results also reinforce the strengths of the PVDT, if the measurement system was valid, a Gage R&R and Process Capability statistics could have been produced to fulfil the final measure and early analyze of the Six Sigma process improvement cycle and add quantifiable information to the process improvement investigation.

The PVDT yields a Multi-Vari study providing a graphical check validating the attention of the project on the measurement system was the correct action, as the Red X was a within-piece problem. It can be taken further to show that if the measurement system was valid, aMulti-Vari would benefit a process improvement initiative narrow down to the root cause of the problem. this is achieved without adding time or cost to the data collection, as the information required is captured in the Gage R&R data and in an objective manner.

The addition of graphical analysis of measurement system variation in the PVDT by taking advantage of an Isoplot^{SM} is important to the diagnosis of measurement system problems. Supplementing a Gage R&R study to give a clear indication where a measurement problem stems from, with the context of statistical information obtained by the Gage R&R. In the case of TRW where tolerances were not obtainable it is of even greater importance to have a graphical test.

These results allowed the process engineers at TRW, to objectively decide to end their follow up of this measurement method. This action was largely based on the compelling evidence provided by the Isoplot^{SM} analysis. In order to improve this non-destructive peel test to a stage where it was viable would require the purchase of a new load cell that is more sensitive to the necessary low loads applied by the Lloyd test machine. Also consideration had to be taken for the implication of the test affecting the characteristic of the weld as this would need a reduction of the limit load being applied. Again pointing towards the need for a more sensitive load cell and ultimately this may not prove effective. The lead engineer on the project from TRW estimated that the efficiency of the PVDT had saved TRW around £10,000 compared to their usual routes of validating an experimental measurement system.

10.4   PIM Game; Real time Practise of Process Improvement Methods

A teaching simulation of a manufacturing process known as the Process Improvement Methodology (PIM) game was introduced. This simulation is used to demonstrate practically a range of Six Sigma and Shainin process improvement tools within the Hybrid DMAIC framework.

This thesis outlines the basic mechanics of the PIM game which are more fully explored in Cox [29]. The need for such a game in the teaching environment was also identified. It is interesting to see how many of the individual process improvement methods and techniques demonstrated by the PIM game have had interactive teaching experiments described in the literature. Indicating there is a need for interactive learning tools to demonstrate process improvement techniques. However the PIM game appears to be the only in classroom tool that takes students through the main stages of the DMAIC cycle using a constant example for students to relate too. Along the way making students implement taught process improvement techniques with the support and supervision of an experienced practitioner. Instead of just lecturing on the techniques and expecting the students to be able to replicate them in industry.

This chapter then went on to test and modify changes to the PIM game proposed in Cox [29]. It was clear from the testing that the proposed changes to the game to include two sided as well as the original single sided specification fault would vastly improve the richness of the learning experience. However there were problems with the game proposed in Cox [29], which required new thinking in order to maintain these principles yet provide clear results when taught techniques were applied to the quality problems the game poses to the students.

Students would now have to understand the subtleties of the two situations and how to apply the require process improvement technique in each case. Part of the authors learning experience was to understand how some of the Shainin tools, which are only described in literature in the single sided distribution case, can be extrapolated and applied to the two sided specification case. Methods such as BOB v WOW and Components Search (chapter 6) were reinterpreted to enable its user to fully understand potential causes and possible best settings of a two sided specification problem.

10.5   Destructive Measurement System Analysis

The final impetus of the thesis was to first propose a new method of validating a destructive measurement system. Then by going back to the previous TRW weld strength problem benchmarking the method against a traditional method of validating a destructive measurement system.

A procedure was developed to test parts with the destructive test equipment as if Homogeneity could be assumed. Using second experimental non-destructive test equipment the size of the variation between consecutive parts could be assessed. If the

results from these two tests showed a strong correlation the assessment of variation could be used to form an estimate of the genuine destructive Gage R&R.

This procedure was tested and benchmarked against using a reference sample at TRW. The scope of this project was to follow on from the previous TRW project were a non-destructive peel test of FET weld strength was proven to be destructive. Therefore a second experimental but destructive pull test was to be validated.

This work made it was possible to develop a reference sample test. From which a Gage R&R of 5.40% was finally determined for the experimental pull test after the development of a reference sample test which did not suffer from the confounding of product variation with that of the repeatability of the test.

The proposed method of using a second experimental non-destructive test equipment to assess the size of the variation between consecutive parts was use to determine a Gage R&R of 18.11% for the destructive pull test. This is a significant improvement on the Gage R&R score of 38% obtained by assuming homogeneity between consecutive parts.

On face value this method would classify the destructive pull test as marginal whereas the reference sample method would classify the test as adequate. This would suggest the reference sample test was the more effective method; however it can be seen in chapter 9 that developing a reference sample piece can be extremely troublesome. With the intricacies of developing a test piece which, in this case, needed to apply loads to the test jig in a very specific manner in order to simulate the test whilst also breaking at a constant load difficult to achieve.

It was also shown that the proposed comparison test could have potentially produced an adequate Gage R&R score of 10% if it had been possible to eliminate reproducibility variation. It was suspected this was created by the randomisation of samples for testing on this occasion consistently resulted in one operator testing stronger samples than the other, rather than one operator testing in a different manner to another.

It can be seen that when there is a second non-destructive test available a conservative estimate of Gage R&R can be determined. In the TRW project the development of a reference sample provided a better estimate of Gage R&R, but this took a lot of work to develop a test piece. There may also be situations where a second non-destructive test method is available to validate a destructive test and it is also not practical to develop a reference test piece.

Appendix A: Minitab Macro

This Appendix includes the code used to produce a macro that automates the plotting of a Gage R&R study, a Provisional Process Capability Study and a Shainin Multi-Vari in Minitab15 from a PVDT sample. This macro was successfully implemented in the described case studies.

```
GMACRO

PVDT

Rowtoc 'Operator 1 Repeat 1' 'Operator 1 Repeat 2' 'Operator
2' 'RESPONSE'.

MVarChart;
  Response 'RESPONSE';
  Factors 'Repeat' 'Piece' 'Batch';
  MuAon;
  MuBon;
  MuCon.

Let 'RESPONSE_1' = 'Operator 1 Repeat 1'

NormTest 'RESPONSE_1'.

Capa 'RESPONSE_1' 4;
  Lspec 'LSL';
  Uspec 'USL';
  Pooled;
  AMR;
  UnBiased;
  OBiased;
  Target 'Target';
  Toler 6;
  Within;
  Overall;
  CStat.

Stack 'Operator 1 Repeat 1' 'Operator 1 Repeat 2' 'Operator 1
Repeat 1' 'Operator 1 Repeat 2' 'EV'.

Stack 'Operator 2' 'Operator 2' 'Operator 1 Repeat 2'
'Operator 1 Repeat 2' 'AV'.

GageRR;
  XBar;
  Parts 'Parts';
  Opers 'Operators';
  Response 'EV';
  Studyvar 5.15;
  SSTUDYVAR 'Repeatability (EV)' 'Reproduceability (AV)'.

GageRR;
  XBar;
  Parts 'Parts';
```

```
  Opers 'Operators';
  Response 'AV';
  Studyvar 5.15;
  SSTUDYVAR 'Reproduceability (AV)' 'Adjusted AV'.

Delete 2:5 'Repeatability (EV)' 'Reproduceability (AV)'
'Adjusted AV'

Let 'Adjusted AV' = IF((( 'Reproduceability (AV)' *
'Reproduceability (AV)' ) - (( 'Repeatability (EV)' *
'Repeatability (EV)' ) / 40 ))>0,SQRT(( 'Reproduceability
(AV)' * 'Reproduceability (AV)' ) - (( 'Repeatability (EV)' *
'Repeatability (EV)' ) / 40 )),0)

Let 'Gage R&R'=SQRT(( 'Repeatability (EV)' * 'Repeatability
(EV)' ) + ( 'Adjusted AV' * 'Adjusted AV' ))

Let '%Gage R&R'=('Gage R&R' / 'Tolerance') * 100

ENDMACRO
```

Bibliography

[1]   M. G. Aboelmaged, "Six Sigma Quality: A Structured Review and Implications for Future Research," *International Journal of Quality & Reliability Management*, vol. 27, no. 3, pp. 268-317, 2010.

[2]   N. R. Senapati, "Six Sigma: Myths and Realities," *International Journal of Quality and Reliability Management*, vol. 21, no. 6, p. 683, 2004.

[3]   M. Harry and R. Schroeder, *Six Sigma:The Breakthrough Management Strategy Revolutionizing The World's Top Corporations*, 1st ed. New York: Doubleday, 2000.

[4]   P. S. Pande, R. P. Neuman, and R. R. Cavenagh, *The Six Sigma Way: How GE, Motorola, and Other Top Companies are Honing Their Performance*. McGraw-Hill Professional, 2000, p. 448.

[5]   J. De Mast, W. A. J. Schippers, R. J. M. M. Does, and E. R. van Den Heuvel, "Steps and Strategies in Process Improvement," *Quality and Reliability Engineering International*, vol. 16, no. 4, pp. 301-311, Jul. 2000.

[6]   M. L. George, D. Rowlands, M. Price, and J. Maxey, *The Lean Six Sigma Pocket Toolbook: A Quick Reference Guide to 70 Tools for Improving Quality and Speed*. McGraw-Hill Professional, 2005, p. 225.

[7]   J. De Mast, "A methodological comparison of three strategies for quality improvement," *International Journal of Quality & Reliability Management*, vol. 21, no. 2, pp. 198-213, 2004.

[8]   D. Shainin and P. D. Shainin, "Better Than Taguchi Orthogonal Tables," *Quality and Reliability Engineering International*, vol. 4, pp. 143-149, 1988.

[9]   P. D. Shainin, "Managing Quality Improvement," in *ASQC Quality Congress Transactions*, 1993, pp. 554-560.

[10]   S. H. Steiner, R. J. Mackay, and J. S. Ramberg, "An Overview of the Shainin System," *Quality Engineering*, vol. 20, no. 1, pp. 6-19, 2007.

[11]   K. R. Bhote, *World Class Quality: Using Design of Experiments to Make It Happen*, 1st ed. AMACOM, 1991.

[12]   K. R. Bhote and A. K. Bhote, *World Class Quality: Using Design of Experiments to Make It Happen*, 2nd ed. AMACOM, 2000.

[13]   K. K. Hockman, "World Class Quality: Using Design of Experiments to Make It Happen by Keki R. Bhote," *Technometrics*, vol. 36, no. 4, pp. 425 - 426, 1994.

[14]   E. R. Ziegel, "World Class Quality by Keki R. Bhote; Adi K. Bhote," *Technometrics*, vol. 43, no. 1, pp. 112-113, 2001.

[15] R. W. Hoerl, "Six Sigma Black Belts: What Do They Need to Know?," *Jounal Of Quality Technology*, vol. 33, no. 4, pp. 391-406, 2001.

[16] International Automotive Sector Group, *QS9000: Measurement Systems Analysis*. 1995.

[17] T. Dasgupta and S. V. S. N. Murthy, "Looking Beyond Audit-Oriented Evaluation of Gauge Repeatability and Reproducibility: A Case Study," *Total Quality Management*, vol. 12, no. 6, pp. 649-655, 2001.

[18] International Automotive Sector Group, *QS9000: Statistical Proces Control*. 1995.

[19] L. A. Seder, "Diagnosis with Diagrams," *Quality Engineering*, vol. 2, no. 4, pp. 505-530, 1990.

[20] J. De Mast, K. C. B. Roes, and R. J. M. M. Does, "The Multi-Vari Chart: A Systematic Approach," *Quality Engineering*, vol. 13, no. 3, pp. 437-447, 2001.

[21] M. Deleryd, "A Pragmatic View on Process Capability Studies," *International Journal of Production Economics*, vol. 58, no. 3, pp. 319–330, 1999.

[22] F. Hunter, "Test Rig Design Validation," University of Durham, 2010.

[23] C. Waddell, "Test Rig Design Validation," University of Durham, 2010.

[24] J. W. Tukey, "A Quick, Compact, Two-Sample Test to Duckworth's Specifications," *Technometrics*, vol. 1, no. 1, pp. 31–48, 1959.

[25] J. Antony, "Spotting the Key Variables using Shainin's Variables Search Design," *Logistics Information Management*, vol. 12, no. 4, pp. 325-331, 1999.

[26] S. H. Steiner, M. Hamada, B. J. G. White, S. Mosesova, and G. Salloum, "A Bubble Mixture Experiment Project for Use in an Advanced Design of Experiments Class," *Journal of Statistics Education*, pp. 1-20, 2007.

[27] C. Chen and H. Roth, *The Big Book of Six Sigma Training Games*. New York: McGraw-Hill Professional, 2005.

[28] S. H. Steiner and R. J. Mackay, "Teaching Process Improvement Using a Virtual Manufacturing Environment," 2008.

[29] S. Cox, "Quality Improvement Methodology," Durham University, 2009.

[30] J. De Mast and A. Trip, "Gage R&R Studies for Destructive Measurements," *Journal of Quality Technology*, vol. 37, no. 1, pp. 40-49, 2005.