Customer-orientated Six Sigma in call centre performance measurement

Article *in* International Journal of Quality & Reliability Management · June 2009

DOI: 10.1108/02656710910966110

Rodney Mcadam
Ulster University

John Davies
University of Salford

International Journal of Quality & Reliability Management

Volume 26, Issue 6, 2009, pp.516-643

Received 21 June 2008
Revised 19 December 2008
Accepted 26 January 2009

Customer-orientated Six Sigma in call centre performance measurement

Rodney McAdam
Ulster Business School, University of Ulster, Belfast, UK
John Davies
Salford Business School, University of Salford, Manchester, UK
Bill Keogh
School of Management and Languages, Heriot-Watt University, Edinburgh, UK, and
Anthony Finnegan
Ulster Business School, University of Ulster, Belfast, UK

Abstract

Purpose – The aim of this paper is to explore the role of Six Sigma performance measurement at both strategic and operational levels within call centres where the definition of Six Sigma is widened to include systems thinking constructs.

Design/methodology/approach – A two-phase methodology is used involving two call centre cases within a call centre group. Phase 1 establishes the need for Six Sigma customer-based measures in addition to internal performance measures and phase 2 studies the implementation of this wider set of Six Sigma performance measures.

Findings – The development and application of Six Sigma performance measures that cover both strategic and operational performance measures lead to a more sustainable approach to business improvement, rather than traditional call centre internal performance measures which may be misleading for the overall performance of the call centre.

Research limitations/implications – The development of the strategic and operational, or double, DMAIC approach offers opportunities for developing wider applications in service contexts using customer-orientated performance measures.

Practical implications – If call centres rely solely on internal performance measures, a misleading picture of call centre performance may be obtained. There is a need to apply Six Sigma to cover both strategic and operational performance measures.

Originality/value – A combined strategic and operational approach to Six Sigma has been developed which enables service-based organisations (call centres) to develop sustainable business improvement.

Keywords Call centres, Six sigma, Performance measures, Strategic evaluation, Operations and production management, Business improvement

Paper type Research paper

International Journal of Quality & Reliability Management
Vol. 26 No. 6, 2009
pp. 516-545

0265-671X
DOI 10.1108/02656710910966110

Introduction

Call centres are often the first point-of-contact for customers and play a key role on whether customers decide to leave or stay with an organisation (Taylor *et al.*, 2003; Conz, 2007; Cleveland, 2007). However, Bellman (2007) indicates that almost 20 per cent of all callers hang up with their issues unresolved. And of those, 68 per cent are at risk of defection. The operating of call centres are normally viewed as a cost to organisations (Conz, 2007), a key factor in organisations outsourcing to other lower wage countries or attempting to make more transactions electronic (Curry and Lyon, 2008). However, as shown by Harney and Jordan, 2008) these initiatives, although positioned as a way to improve customer experience, are mainly internally focussed, looking at cost and performance. The result of these measures being incorporated into scorecards as targets means that managers are now focussed on improving the output and discussing issues such as "how can abandoned calls numbers be reduced to 5 per cent?", rather than focussing on the input - "what are customers calling about and how well are their calls first time?". This focus has meant there is also an absence of leading performance measures to better position the call centres to react to and improve further the customer experience.

Call centres, although being service organisations, are data centric in that they record data on performance over long periods of time. The availability of such data along with the need for business improvement offers the possibility of applying business improvement methods that have been developed in other sectors. One such approach is that of Six Sigma where there is a reciprocal challenge first for call centres to avail of the full range of Six Sigma based improvement, and second, for Six Sigma to be adapted to address strategic and operational issues in a service based environment.

The aim of this paper is to explore the role of Six Sigma as a performance management system at both strategic and operational levels within call centres where the definition of Six Sigma is widened to include systems thinking constructs. This aim reflects the challenge facing call centres which is how to deliver an excellent customer experience while simultaneously reducing costs and by understanding the true demand from customers and determining call centres capability to deal with that demand, looking at improvement in a holistic way.

Call centre operations

The initial call centres were in-house operations in larger organisations (Koh *et al.*, 2005). The advantage of structuring in this way is that of a coherent department focused on telephone services. Moreover, there is the cost benefit of having more calls handled by fewer people (Hart *et al.*, 2005). However, there are problems with this approach as shown by Koh *et al.* (2005). First, it led to stretch targets by way of calls per hour. Second, it led to a high turnover in staff as people became demotivated by the pressurised atmosphere. The Contact Babel (2006) UK contact centre review points out that almost all studies over the previous four years have reported average staff attrition rates at between 15 per cent and 25 per cent. In 2006 the figure was 23 per cent which was the fourth annual increase in a row which results in increased costs of recruiting and training linked to the high staff turnover. Seddon (2001) found that resolution at the first point-of-contact in call centres varied but was never higher than 65 per cent and was as low as 21 per cent in financial service organisations. Gettys (2007) determined that a Lean Six Sigma expert carrying out basic process analysis at a financial services call centre found that: the majority of calls that could not be resolved on the first call required some research by the service representatives; the service representatives were primarily judged on whether they were available to answer, limiting the time they could devote to other issues, and customers whose inquiries were not answered within a few days would call back. This increased the call volume, inflated the numbers of calls that could not be resolved on the first call, and led to multiple entries in the computer system for the same problem.

This focus on performance measures of quantity, how many calls and how quickly they are answered, has resulted in the majority of call centres failing to learn about the customer or to establish relations (Curry and Lyon, 2008; Hart *et al.*, 2005). Moreover, an emphasis on keeping call times as brief as possible could actually cause the agent, at best, to sound impersonal and unsympathetic to the customer. At worst the customers" enquiry will not be fully resolved (Seddon, 2001) leading to further calls and possible damage to the company's reputation with the customer. Nevertheless, this early approach to call centres generated very substantial efficiency and cost gains for companies (Taylor *et al.*, 2003).

From the early 1990s onwards Customer Relationship Management (CRM) has become synonymous with call centre operations. The integration of information telephony systems, designed to provide advisors with information they need to service the customer, has led many call centres organisations to apply CRM. However, Seddon (2005) states that call centres are still designed in much the same way as the mass-production manufacturing factories over the last century with a history of worker alienation, high staff turnover and low morale with the decision making being separated from the work, the last point having it's roots within Taylor's scientific management principles first published in 1911 (Curry and Lyon, 2008).

Technology has come a long way within call centres over the last 20 years. From telephony switch boards manned by people with separate IT systems, to today's environment where Telephony, IT delivery systems and reporting suites are fully integrated giving the call centre a complete view of the customers journey through these systems. Bellman (2007) points out there are many and varied tools that organisations can utilise:

> This diverse toolkit includes interactive voice response (IVR), knowledge management systems, agent scripting, presence and instant messaging, real-time speech analytics, and training and policy.

Difficulties arise when this technology has been designed to meet internal operational performance measures without fully understanding the customer needs and/or the support functions needed (Bellman, 2007). Curry and Lyon (2008) and Koh *et al.* (2005) determine that call centres face the challenges of providing an outstanding customer experience while at the same time improving productivity.

Six Sigma in services and call centres

From a sectoral standpoint the Six Sigma literature is predominantly manufacturing based, with mass manufacturing being the basis for most studies (e.g. McAdam and Lafferty, 2004). However, there is evidence in both the academic and practitioner literatures that Six Sigma developments and applications in other organisational sectors and functions are growing rapidly (e.g. Antony *et al.*, 2008; Chakrabarty and Tan, 2007), which is an indication that the Six Sigma discourse is deepening (de de Koning and de Mast, 2006; McAdam and Lafferty, 2004). However, there is a need for further research to support the descriptive based claims in service sector studies (Proudlove *et al.*, 2008; McAdam and Lafferty, 2004) and the more involved and complex people interactions as opposed to machine dominance (Sehwail and DeYong, 2003). Antony *et al.* (2008) and

Chakrabarty and Tan's (2007) review of Six Sigma applications in services concludes that the development is slow but increasing in terms of structure and more in-depth applications and that more research in this area is needed.

A key development of service based Six Sigma is in the data centric service organisations involving long run data streams. These applications are found mainly in the healthcare sector, both public and private due to Government emphasis on health sector reform; (see for example, Peltokorpi and Kujala, 2006 – hip replacements; Morgan and Cooper, 2004 – Rehab; Revere and Black, 2003 – patient care). These top down agent based applications of Six Sigma focus on the operational methodology of Six Sigma rather than wider strategic interpretations. Hence, there is an emphasis on applying the key methodology within Six Sigma, namely Define, Measure, Analyse, Implement, Control (DMAIC) and the key success factors that need to be addressed to implement DMAIC in specific service based contexts (Antony *et al.*, 2008).

This emphasis on translating Six Sigma methodology from mass manufacturing to that of a service based context with attendant consideration of contextual key success factors reflects a tendency in quality and operations literature to apply business improvement approaches across sectors at a methodological level without considering the wider strategic assumptions and implications that lie behind such approaches (de Koning and Mast, 2006). It is therefore suggested that an inquiry into Six Sigma in service based organisations, and call centres in particular, should cover both strategic and operational aspects of Six Sigma, which will expose an organisation to the full range and depth of the Six Sigma philosophy, rather than using limited applications at lower operational levels.

The need for strategic and operational developments of Six Sigma is stressed by Nonthaleerak and Hendry (2008) who critique the DMAIC formula based on a multiple case analysis ($n = 9$). Their study suggests that the Define and Control steps within operational applications of Six Sigma have limitations in that the Define step and its associated criteria for selecting Six Sigma projects may focus on lower level problems as opposed to strategic customer based opportunities. Similarly, the Control step may fail to create strategically sustainable gains in strategic projects due to their cross functional nature which precludes attribution of ownership. Similarly, Friday-Stroud and Butterfield (2007) suggest the need to incorporate strategy and decision making constructs to develop a more robust or expanded DMAIC.

In attempting to address these issues and devise a framework for Six Sigma within call centres a systems thinking approach has been used to integrate both strategic and operational levels of Six Sigma. Seddon (2001) proposes a six stage approach called the Vanguard Model when applying systems thinking in service organisations. This model addresses the wider and more strategic issues within an organisation when attempting to apply new knowledge, namely Six Sigma in the current study. Amelsberg (2002) suggests that systems thinking concepts can be used to integrate strategic and operational aspects of Six Sigma by using a double DMAIC approach. The first DMAIC is used at a strategic level and is an interpretation of DMAIC from a strategic systems perspective. The resultant outcomes are then used to start the second or operational level DMAIC. Table I illustrates how the strategic level DMAIC can be interpreted using systems thinking concepts (key stages in the Vanguard model) (columns 1 and 2) leading to the operational DMAIC (column 3). Friday-Stroud and Sutterfield (2007) have used a similar approach in suggesting steps in strategic

Strategic level Six Sigma service issues	Systems thinking issues	Operational level Six Sigma methodology
Define: key strategic objectives to achieve breakthrough improvements	Purpose: What is the purpose of the call centre service from the customer's perspective rather than from an internal cost or functional view?	Define – define Six Sigma projects consistent with the strategic objectives
Measure: the entire business systems that support the strategic objectives	Identify what the true nature of demand is from the customer's external perspective, not from the call centres internal view	Measure – measure performance of the Six Sigma projects at an operational level
Analyse: determine gaps in the systems performance measures and benchmarks	Measuring the call centre's capability to deal with this demand at the first point-of-contact for the customer and what are the barriers that prohibit first contact resolution	Analyse – analyse project performance in relation to operational goals
Improve: focus on improving systems elements to achieve customer performance goals	Flow: map the flow of the work as an end-to-end process based on customer demand Value: understand how the value sought by the customer will help the call centre in designing to increase its capability to meet this demand and identify new offerings and opportunities Waste: identifying how much waste is in the system and more importantly how it flows through the system, to enable the call centre to increase value for the customer	Improve – improve the key service through internal performance measures
Control: characteristics that are critical to value and standardise and integrate with the call centres strategic plans	Focuses on identifying what "systems conditions" impede meeting the customer's demand, challenging "command and control" thinking, and embedding a strategic Six Sigma philosophy as the corner-stone for achieving organisational excellence	Control: establish sustainability of improvements in operational processes and measures with continuous improvement

Table I.
Strategic and operational DMAIC

management a precursor to applying DMAIC at an operational level. However, the systems thinking approach, as adopted in Table I, has advantages in that rather than simply translating strategy to operational level DMAIC, it uses Six Sigma concepts at a strategic level which is then integrated with Six Sigma at operational level. This approach recognises Six Sigma as a business improvement philosophy at both strategic and operational levels within an organisation.

This strategic and operational approach to Six Sigma, using systems thinking concepts, challenges service organisations such as call centres to both deliver customer requirements and reduce operating costs (Amelsberg, 2002). It questions the sole use of internal performance measures and targets that managers use to drive and assess performance in call centres. Hence, there is a need to integrate strategic and operational level decision making and measurement in the approach to Six Sigma as shown in Table I.

Research questions and methodology

The research issues arising from the aims of the study and the literature discussion were expressed as "how and "what" type research questions as suggested by Yin (2003):

RQ1. How can Six Sigma be used to improve call centre performance at both strategic and operational levels through the identification of both customer and internal measures?

RQ2. In what way can the Six Sigma DMAIC methodology be applied to cover both strategic and operational improvement in call centres; especially in regard to the Define and Control steps?

Both Yin (2003) and Eisenhardt (1989) suggest that an interpretative research stance is more appropriate to these types of research questions to enable in-depth inquiry. Eisenhardt (1989) shows that reflection and comparison with existing theory or external comparisons can help in bringing objectivity to the process (i.e. Phase 1). Multiple sources of data are embraced and engaged in a recursive sense making process (Phase 2). The chosen research methodology was that of case studies which is suited to the interpretive research approach (Yin, 2003).

Two case studies from within a larger organisation or group of call centres, were chosen for the study. These call centres were mature in nature with established processes and measurement functionality. The case studies covered were the helpdesk call centre referred to as call centre 1 (phase 1) and the Repair call centre (call centre 2 – phase 2).

The call centres' were each set a plethora of new internal targets by the parent organisation, including Abandoned Call Rates (ABN), sales achieved, call handling time, and per cent of calls answered within 15 seconds. Resource teams were also split to manage these new teams, schedule their attendance and report on their performance. The routes into the call centre were also changed to ensure customers were directed to the most appropriate team, this included 0800 numbers and Integrated Voice Response Systems (IVRs). Although the decision to restructure in this way was mainly to help better serve the customer, there was no customer demand analysis carried out to support or challenge if the new structure would increase the capability of BT Ireland to serve customers better.

There was a two-phase approach to the research. Phase 1 probed the use of performance measures in call centre 1 to determine an appropriate set of performance measure for Six Sigma application beyond that of the existing operational measures used within call centre 1. The second phase used this expanded set of measures to apply strategic and operational Six Sigma to call centre 2.

A multi-layered research approach using a combination of primary and secondary research was used in both phases, including participant-observer methodology in phase 2 after the manner suggested by Remenyi *et al.* (1998). The framework analysis was the key steps shown within Table I, which cover the double DMAIC or combined strategic and operational approach to Six Sigma from a systems thinking approach.

The number and frequency of meetings, interviews and focus groups are shown in Table II. Internal secondary data was gathered through interviews, phone calls, system interrogation and accessing company records.

Instrument	Purpose	Outcome	Number
Phase 1 – Case analysis and establishment of strategic and operational measures for Six Sigma – case study 1			
Interviews	To understand change within the case organisation	Confirmation that BPR and TQM are both utilised in BT Ireland	10
	To understand the current structure of call centres in case organisation	Current structure defined and outlined within research	
Focus groups	To review case study data gathered and discuss next area of analysis	Data analysis completed for case study	8
Workshops	To share case study analysis findings with call centre management and their teams and to gain approval for further analysis	Data shared and published and all approval sought granted	4
Phase 2 - Proof of concept – case study 2			
Interviews	To check progress on DMAIC steps with key managers	Achievements and problems	10
Focus groups	To review proof of concept data gathered and discuss next area of analysis	Data analysis completed for proof of concept	30
Workshops	To share proof of concept analysis findings with call centre management and their teams and to gain approval for further analysis	Data shared and publish and all approval sought granted	8

Table II.
Research methods

The primary data collection process required for the case study and proof of concept was carried out by a combination of both the researchers and the intervention team. One of the researchers acted in a participant observer role as part of his role of Business Improvement Manager and supported the intervention team in analysing and presenting the findings. Internal secondary data consisted of current performance on calls offered; call handling time, abandon calls and percentage of calls. This data was analysed using Six Sigma software to produce data run charts.

Results and discussion

Phase 1 – Case analysis and establishment of strategic and operational measures for Six Sigma – call centre 1

The type and frequency of calls coming into the call centre were measured. The call type analysis showed that three distinct types of calls were received:

(1) value;

(2) help; and

(3) failure.

A template was established that enabled the frequency of each type of call to be measured – 1,015 calls were captured and categorised in this phase (Figure 1).

The analysis showed that Value made up 20 per cent of calls, Help was 60 per cent and Failure was 20 per cent. From this data the researcher was able to calculate the capability of call centre 1 to handle their customer demand at the first point-of-contact as 68.5 per cent. This result is consistent with Seddon's (2001) findings of 60 per cent to

VALUE	201	20%
HELP	606	60%
FAILURE	208	20%
TOTAL	1,015	

Resolved/Dealt with on line
Passed activity/contact to other unit to resolve/deal
Advised customer to speak to another organisation
No order/request placed

Contact Types	You were able to resolve on line	You raised an order	You raised an activity for duplicate bill, itemisation, recon. etc.	You raised an open issue and placed in a queue	You had to transfer to another BT department or give alt. number	You had to advise customer to speak to another organisation	Customer did not place an order/request	
VALUE								
I want a product/service	20	113	4	4	27	1	32	201
HELP								
I want to do something/ I want you to do something for me	109	66	42	18	58	7		300
I need information	126	9	4	4	22	3		168
I don't understand	54	4	0	11	7	1		77
I have a problem using/ with a product/service	10	4	0	2	40	5		61
	299	83	46	35	127	16		606
FAILURE								
I don't agree with the charges/bill	30	5	3	11	2	0		51
There's a problem with BT plant/Civil issues	2			1	1			4
You haven't done what you said you would	18	11	9	45	21	1		105
You've done something............. I didn't ask you to do/you didn't tell me about/you shouldn't have done	11	2	5	11	6			35
I'm unhappy with the level of service	3	0	1	3	6			13
	64	18	18	71	36	1		208

Figure 1.
Analysis of value, help and failure call categories

70 per cent in service organisations. Further capability analysis showed that although Value and Failure calls made up 20 per cent each of the overall calls total, advisors were almost twice as likely to resolve a Value call (84 per cent capability) at the first-point-of-contact versus 48 per cent capability on a Failure call (Table III).

Not getting things right for the customer at the first point-of-contact resulted in almost 30 per cent of the calls received being passed to other parts of the organisation to get resolved, meaning they have to be investigated by more than one advisor and necessitating a call back or letter to be generated to ensure the customer obtained closure on their issue, hence requiring further resource and thus increasing costs (Hart *et al.*, 2005).

	Dealt with/resolved online (%)	Activity/contact passed to other division to deal (%)	Customer advised to speak to other organisation (%)
Value	84.0	15.5	0.5
Help	70.5	27.0	2.5
Failure	48.0	51.5	0.5
Total	68.5	29.5	2.0

Table III.
Call centre 1 capability to deal at first point-of-contact analysis

The handling cost of an online call cost the call centre £3. This is similar to the costs presented by Contact Babel UK call centre review (Contact Babel, 2006). When passed to offline teams customer's enquiries were touched at least 2.5 times, increasing that cost to £10 per call. The researchers also tracked and analysed the standard call centre measures that were in existence, namely call handling Time (CHT), abandoned call rates (ABN per cent) and percentage of calls answered in 15 seconds (PCA15) which are standard internalised call centre performance measures (Curry and Lyon, 2008).

Using Six Sigma analysis involving Statistical Process Control (SPC) run charts these measures and their variation were analysed using weekly intervals performed over a four-year period. The team identified that call handling Time (CHT), abandoned call rates (ABN per cent) and percentage of calls answered in 15 seconds (PCA15) had all improved from August 2006. Call handling time had improved from a mean of 357 seconds to a mean of 285 seconds (Figure 2).

Abandoned call rates had reduced from a mean of 13.8 per cent to a mean of 4.8 per cent (Figure 3), another excellent improvement in performance.

Percentage of calls answered within 15 seconds increased from a mean of 32.6 per cent to 47.1 per cent (Figure 4), meaning that 47.1 per cent of customer where having their calls answered within 15 seconds of calling. This improvement also has a direct impact on ABN call rates as customer are more likely to hang up if the are being held in a queue for long periods.

The strategic approach to Six Sigma (Table I) indicates that a customer perspective on performance measures or customer central to quality issues (CTQs) is also needed (Friday-Stroud and Sutterfield, 2007). Thus, in. line with the actual internal measure improvements (Figures 1-4) the researchers and the intervention team, focussing on customer demand, identified that call volumes i.e. demand, had also been increasing. This is highlighted with notes in Figure 5. The mean increased from 13,776 calls per week to 18,218 calls per week, an increase of 24.5 per cent. Given that this trend was going against a declining customer base, and improvements in the internal call centre measures, thus the reasons were probed. The results were shared with a group of advisors, coaches and managers from the call centre 1 teams at a workshop where the researcher facilitated a cause and effect analysis to help the group determine what had caused these changes. It was discovered that five new people had been added and the date they started working coincided with the change in the call centre's performance. Further workshops held with the call centre senior management team revealed that the five new people had been sourced from an employment agency and were brought in to help deal with an increase in call volumes and to help the centre meet it's abandoned call target of <5 per cent. It was envisaged that this additional resource would also ensure that customers did not hang up and were answered promptly. These advisors were known as "call grabbers". Their role was to answer calls politely, write the customer's information and request on a sheet that had been designed for the advisor. Customers were then advised that they would be contact them with 24-48 hours to have their request dealt with. The records were gathered up by an experienced adviser throughout the day and distributed out to other experienced advisors with a view that these customers would be called back during periods of low call volumes. The impact of the call grabbers had the desired affect on the internal call centre measures as illustrated in Figures 1-4. The researchers however, decided to analyse the impact that call grabbers had on the customer and the business in line with the strategic and

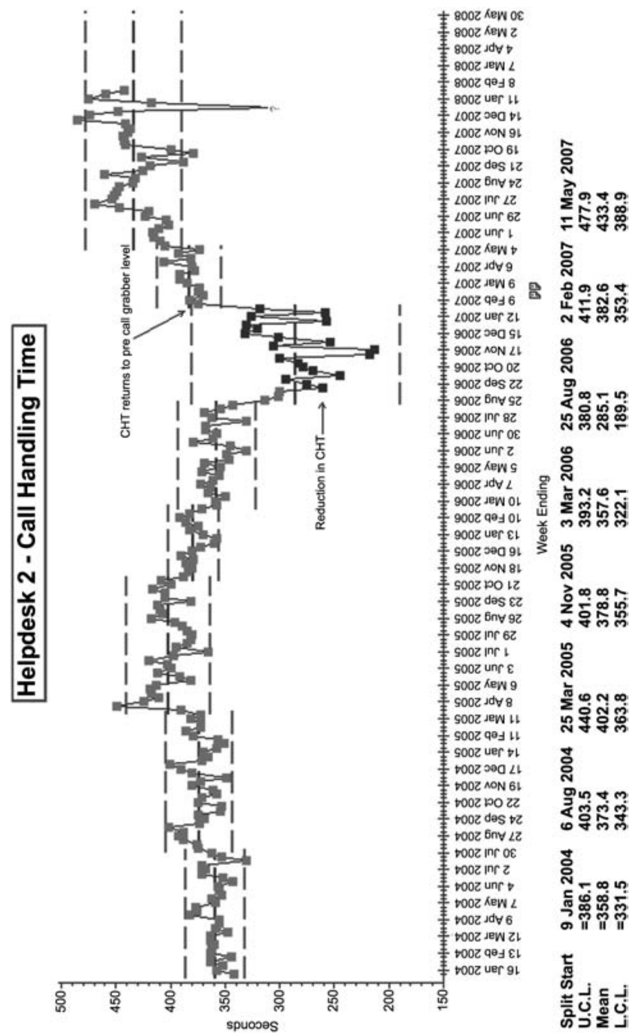

Figure 2.
Call centre 1 – call handling performance from January 2004 to January 2008

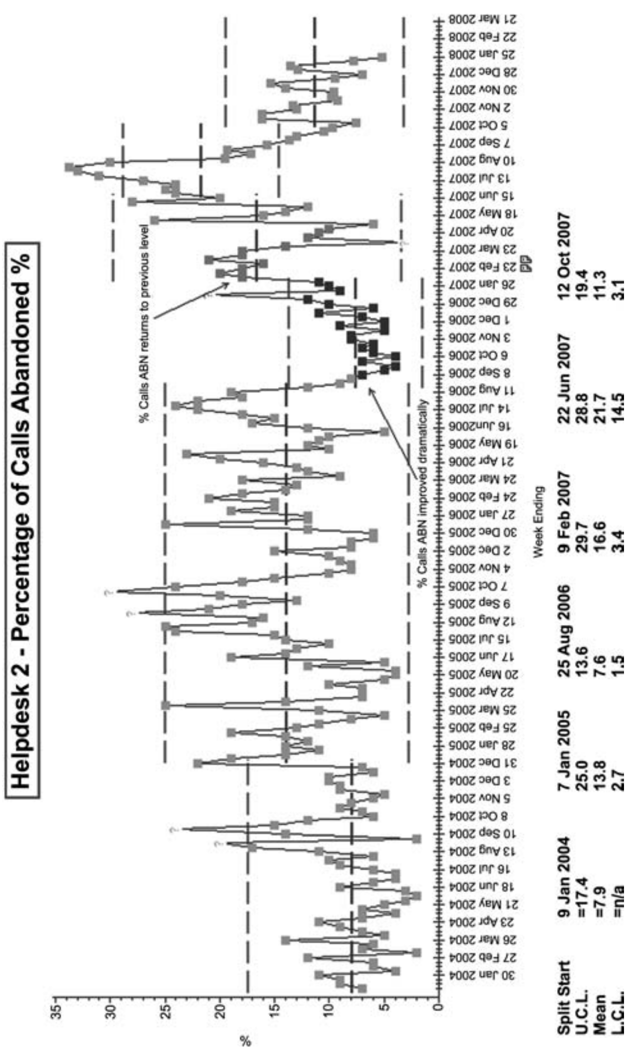

Figure 3.
Call centre 1 – call abandoned rates per cent from January 2004 to January 2008

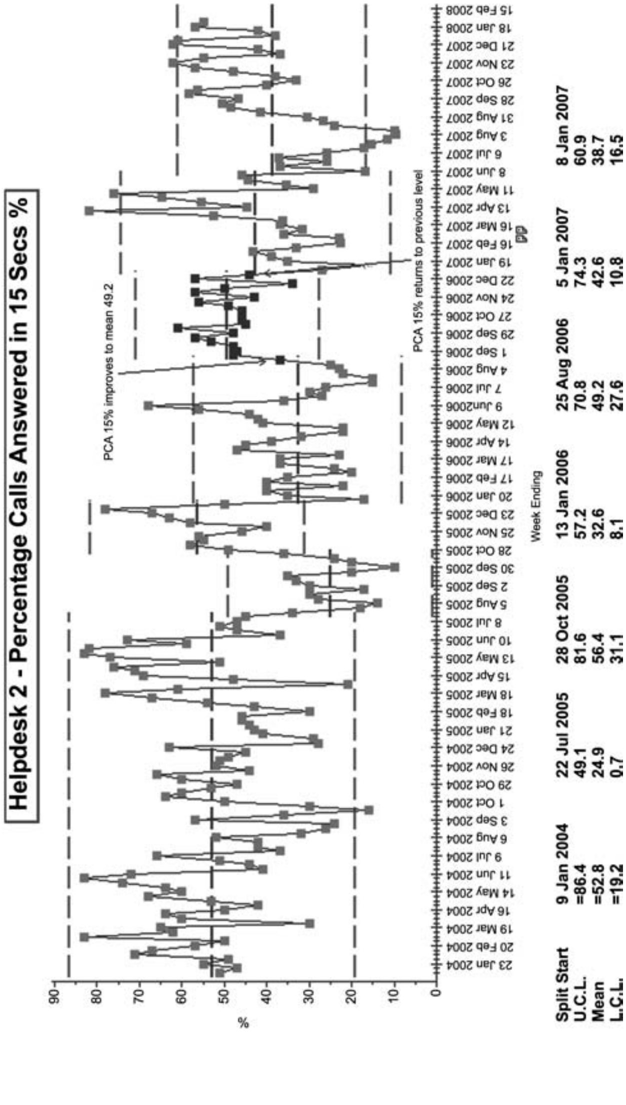

Figure 4.
Call centre 1 – Percentage calls answered in 15 seconds from January 2004 to January 2008

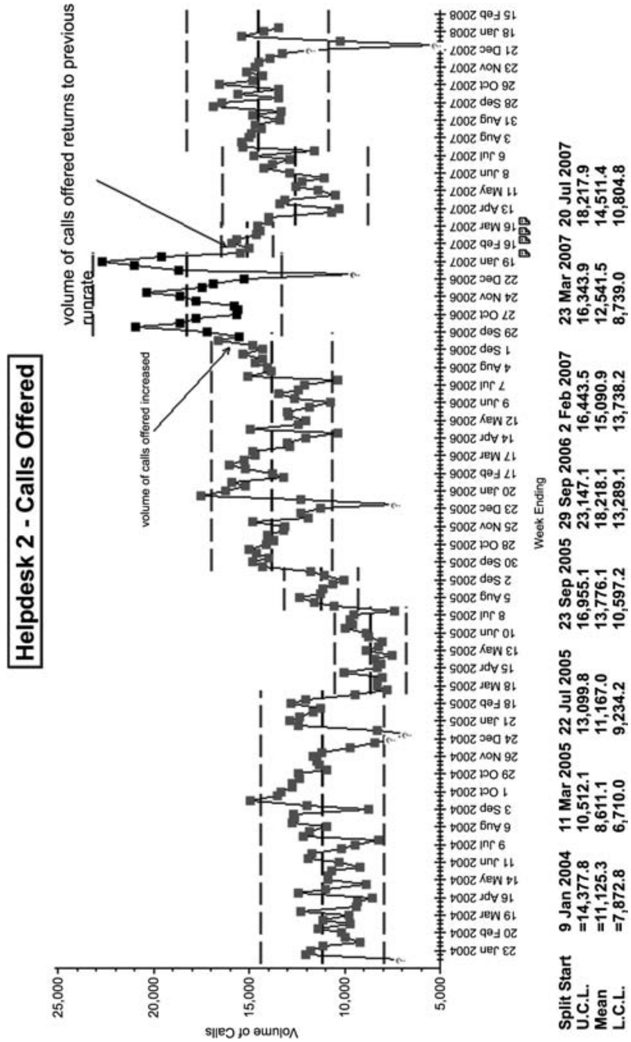

Figure 5.
Call centre 1 – Calls offered from January 2004 to January 2008

systems approach to Six Sigma (Table I). In order to review and share the findings of the analysis the intervention team organised and ran a weekly focus group to analyse the impact on customers and the organisation. Over a three-week period 718 calls that had been taken by the call grabbers were analysed. The analysis showed that advisers were unable to contact 35 per cent of customers when they called them. Added to this 22 per cent of customers called back into the call centre before the experienced advisers had the opportunity to call the customers back (Figure 6). When calculated it was clear to see how the call grabbing process was adding additional calls into call centre 1, while making the jobs of advisers much more difficult (despite the promising picture shown in the internal measures, Figures 1-4).

Each call grabber was answering up to 75 calls per day, over five days, a total of up to 1,875 calls per week. Thus, 805 call backs per week were successful, 412 calls per week repeats (customer called back before the call centre could call the customer), 69 per cent of these customers called back within the first 24 hours of their initial call, the rest after 24 hours, 656 outbound calls were made by experienced advisers, but the customer was not spoken to, thus these customers were likely to call back into the call centre.

The intervention team shared the findings at a focus group attended by advisers, coaches and managers. The attendees were then encouraged to list the qualitative, less measurable affects that this focus on only improving internal measures was having on their day-to-day operations; the findings were grouped as follows:

- *The impact on the management team:* overall management responsibility of additional people; monitoring timekeeping and sick leave; customer escalations from the Call Grabbers; budget expenditure; delays in cancellations or start order; poor customer experience; standard of call grabbers – a number of dismissals – involving coach observing calls; collection/counting and increased pressure from senior management to meet targets.
- *The impact on the advisers and coaches:* Customer dissatisfaction about delay of call back – making call more difficult to handle and leading to escalations; additional pressure from managers to answer and deal with calls faster than

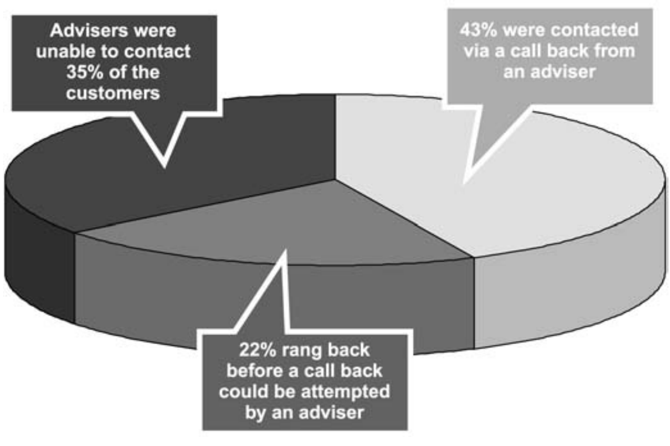

Figure 6.
Analysis of Call Grabber impact on call centre 1 (718 calls)

normal to meet targets; additional pressure to work overtime in order to help out; increased pressure contributing to increased sick leave.

- *The impact on the business:* increased costs of paying the agency for hiring the call grabbers; increased overtime costs and increased customer dissatisfaction.

The final focus group was attended by the senior management team of both call centre 1 and call centre 2 and a number of advisers, coaches and managers from within the Group and was facilitated by the researchers. Based on the quantitative data presented, the team agreed that call grabbing in the call centre, although improving the internal measures, was adding no value to customers and was increasing calls in and costs. The decision was then taken by senior management to cease call grabbing activities immediately (January 2007). The impact of this has can be seen in Figures 1-4, where call handling time, abandoned call rates, percentage of calls answered in 15 seconds and volume of calls offered returned to their pre call grabber average. This return to normal state further emphasised that the introduction of call grabbing to improve internally focused measures had impacted negatively on both costs to the organisation and customers.

The findings support the views expressed by Reynolds quoted in Bellman (2007) that focusing on quantitative performance related measures such as CHT leads to the desire to finish calls quickly rather than resolving customers' issues and hence increasing costs. Phase 2 of the research, the proof of concept, was initiated, to determine if focusing on delivering what matters to customers reduces costs and improves internal quantitative measures, as suggested in the approach to Six Sigma shown in Table I (Antony *et al.*, 2008; McAdam and Lafferty, 2004).

Phase 2: Proof of concept at project level – call centre 2

Building on the knowledge gained in phase 1 the researchers focused on applying the double DMAIC approach (Table I) the repair team (call centre 2) which was responsible for answering calls from customers who have difficulty using their phone service. The volume of calls can be affected by bad weather conditions which increase the likelihood of faults occurring in the network. The call responder's role is to diagnose the customer's problem and send it to the appropriate team to resolve. At this point, a call diversion service is offered where customers can have their calls diverted to a mobile or other number free of charge, ensuring they do not miss any calls. Once the fault is cleared the call diversion has to be manually cancelled. Where no fault is detected the customer is advised to check their own equipment and call back in. The repair team manage a range of offline queues, where customer faults are queued until the customer checks their own equipment or the adviser makes a call-back. Before the proof of concept (phase 2) commenced the repair team did not track the amount of work that flowed through these queues. Their measurement of performance was restricted to that of call quantity and faults raised. The proof of concept followed the double DMAIC approach (Table I) and is discussed under each step as follows.

Define. Working with the intervention team, the repair advisers determined the customer issues and the effect on the organisation as a whole, followed by the purpose of their roles in relation to these issues as suggested by Conz (2007). The team developed a purpose statement that reflected what their role was within the organisation as a whole and the call centre in particular for delivering what mattered to customers and the organisation. From the output of the focus group the team defined and accepted a new purpose statement which gave them something to aspire to while allowing them to devise measures that related to the purpose and would determine if the repair team were achieving their purpose (Curry and Lyon, 2008). The statement agreed was:

> To deliver a World class customer experience by . . . resolving customer enquiries at first point-of-contact in an effective, timely and cost efficient manner . . . and by . . . identifying the customer's problem and where necessary, routing it to the appropriate dealer group to expedite solutions first time.

The team organised a focus group and used their collective knowledge to identify the key drivers customers would see as a good experience. They also listed the issues that were important to the organisation. The key drivers for delivering what matters to customers are as follows:

(1) *Customer drivers*:
- on time;
- right 1st time;
- deal with it at the 1st point-of-contact; and
- give it to the person who can do it.

(2) *Business drivers*:
- cannot improve at all costs;
- we are effective and efficient;
- on time; and
- right 1st time.

It was envisaged that the new customer focused measures devised would enable call centre 2 to measure their ability to resolve more enquires at the first point-of-contact, and to answer calls in a timelier manner.

Measure and analyse. In order to determine the customer demand within the repair call centre (No. 1) the team had first to develop a typology for the type of calls that came to the team. The researchers decided to approach this phase of analysis in a way that looked at the demand from the customer's perspective (Koh *et al.*, 2005; Harney and Jordan, 2008). A group of five advisers volunteered to sit beside a colleague and listen to their calls. The advisers then wrote down exactly what customers said. This information was reviewed at the end of each day. A list of call types was produced in three specific categories. To facilitate gathering the frequency of each type and to ensure no ambiguity arose relating to what each category meant, the team provided an explanation for each:

(1) *Help* – These calls are from customers who require assistance in using, or information about, products and services or need help in detecting and resolving a problem they are experiencing.

(2) *Repair* – A call only becomes a Repair call type when a fault has been identified as the responsibility of the organisation and action is needed to resolve the issue.

(3) *Failure* – These are calls from customers who have failed to receive an appropriate service from the call centre (e.g. customer requests not carried out;

having to chase the progress of an enquiry/problem) – in other words failing to get it right first time. This type of call can be due to failure within our Help or Repair processes.

The type analysis was then shared with the whole repair team who worked together to complete the frequency analysis. A template was then constructed using the call types in order to measure the frequency of the call types. The repair team listened to 1,004 calls over a three-week period and compiled and reviewed the data at the end of each day. The high level output can be viewed in Table IV. Of the 1,004 calls, 14 per cent were classed as inappropriate as customers had either called the wrong number or had to be transferred to another call centre within the Group to have their enquiries dealt with, a symptom of IVR mis-operation. The team decided to exclude these calls from their scope of focus as they felt they had little or no control of where they received their calls from. Therefore, they focused on the 864 calls that were appropriate to their function.

A workshop was held to discuss the results with the repair team. The team were surprised that only 27 per cent of the 864 received were true repair calls. On further investigation the researcher found that the repair team raised a fault for each Help and Repair call received. They also raised a fault on Failure calls, where the original fault had been closed but where the customer had confirmed their fault had not been resolved. Taking the Help calls only as one example the team concluded that they were raising faults on 51 per cent of their calls although no fault existed. It was also confirmed with the repair management team that this was the process advisors followed as "this was how we determine how much work they get through" (i.e. an internal measure that did not address the effect at strategic and systems levels) as suggested by Friday-Stroud and Sutterfield (2007).

Capability of response is the ability of the team to deal with the call at the first point-of-contact or pass it directly to someone who can. The analysis showed that on repair calls the team detected and despatched 84 per cent of their work directly to someone who could resolve it for the customer (Table V). Even though repair calls made up only 27 per cent of their overall total, the repair team were quoting possible time related charges to all customers who fell into the help or repair category, 78 per cent in total. The team questioned the rationale behind quoting time related charges on all these calls when only 16 per cent had no fault detected and where fault may be within the customer's equipment. They identified this as a key issue in delivering what matters to customers.

Frequency analysis	No	%
Total calls analysed	1,004	
Wrong numbers	30	
Transfers to other BT departments	110	
Inappropriate calls	140	
Help	440	51
Repair	235	27
Failure	189	22
Appropriate calls	864	

Table IV.
Call centre 2 repair frequency results

On Help calls (Table VI) the team passed on 13 per cent as faults offline as they were unable to conduct tests (UTT), while they resolved 34 per cent at the first point-of-contact after testing for the customer. A total of 53 per cent were tested and placed into the offline queues awaiting customers to test their own equipment. Customers were encouraged to test their equipment and call back into the team or advisors who subsequently agreed a time to call these customers back to confirm if they had found a problem in their own equipment or if the fault still existed. The call back process can prove difficult as, although a time is agreed with customer, call backs are normally not carried out during periods of high call volumes as the team had a target of <5 per cent abandoned calls. The team therefore identified this as a key problem within their process of delivering what matters to customers.

The team determined there were two types of failure demand. Type 1 related to customers calling back before action had been taken to resolve their problem, 129 calls (Table VII), while type 2 related to faults where action had been taken, but the customer still had a problem, 60 calls (Tables VII and VIII). By looking at the data in

	Fault detected and despatched	No fault or problem detected but customer wants a visit	Total
Repair	197	38	235
%	84	16	100

Table V.
Repair capability analysis

	Fault passed to off line for UTT	Help given/contact resolved – no call back required	Help given – call back required	Total
Help	57	149	234	440
%	13	34	53	100

Table VI.
Help capability analysis

	Customer rang before the ERT/APPT	Customer rang on the day of the ERT/APPT	Customer rang after the day of the ERT/APPT	No ERT or APPT assigned yet	Fault report closed	Total
Failure	50	40	30	6	3	129
%	39	31	23	5	2	

Table VII.
Failure (type 1) capability analysis

	I was able to assist/resolve the issue	I had to get someone else to help resolve issue	I had to transfer the customer to another dept	Total
Failure	43	11	6	60
%	72	18	10	100

Table VIII.
Failure (type 2) capability analysis

this way the team were able to determine if the failure occurred during the customers' original issues, or if the organisation had dealt with the customers' issues, but not fully resolved it to the customers' satisfaction.

The analysis showed that in 70 per cent of calls relating to the customers' original problem, the customer called back within the estimated response time (ERT) or the appointment time (APPT) they had been given. They also found that a further 23 per cent of customers called the day after the ERT or APPT. From this analysis they concluded that the customer did not feel confident that their issue would be resolved when the organisation said and/or they had not been communicated to by the organisation to advise the commitment date could not be met. The team identified this as a key issue in delivering what mattered to their customers.

Where the organisation had dealt with the customers' issue but had not resolved it fully (Table VIII), the team found this was down to mainly one issue, namely call diversion with 15 per cent of all failure calls due to the call diversion service not being set up correctly or not being cancelled when the customers fault had been cleared. This was something that could be resolved by the adviser quite easily as the 72 per cent resolution shows in Table VIII, it was however unnecessary failure. The team identified this as a key issue in delivering what matters to customers.

Therefore, by following the double DMAIC approach as shown in Table I the team had now a full understanding of the type and frequency of their customer demand and also their capability to respond to that demand. They also had a list of key issues that affected how they delivered what matters to their customers (Curry and Lyon, 2008). These are listed in the Improve stage of DMAIC that follows (Table IX).

Improve. These findings gave the team their focus for improvement within the Implement step within DMAIC. In order to complete the end-to-end view from the

Issue	Action to resolve
Time-related charges were being quoted on all repair and help calls even though an engineer visit may not be required	Only quote TRCs when a visit has been arranged, or when a customer refuses to check their own equipment
The current process actively encouraged customers to call back into the team whether a fault was detected or not	Advise customers to call back only when a fault still exists after they have checked their equipment
Faults were placed in the offline queues in order for a call back to be made to the customer, a waste of valuable adviser resource	Only raise a fault and place in the call back queue when the customer has specifically requested a call back. As above, ask the customer to call back only when a fault still exists after they have checked their equipment
70 per cent of customers called into the team before or on the day their fault was due to be repaired, suggesting customers were not confident their issue would not be dealt with	Only raise a fault when a hard fault has been detected and be specific when advising of the date of repair. This will ensure that false expectations are not set with customers
The call diversion process contributed 15 per cent to the overall failure demand	Share call diversion analysis with the team responsible for setting up and cancelling this service and work with them to remove this unnecessary waste

Table IX.
Repair action plan to deliver what matters to customers

customers' perspective the team set about mapped the flow of their work, as suggested by Morgan and Cooper (2004), based on the following question:

- How good are we at delivering what matters to our customers?

The researcher working with the repair team in his role as business improvement manager facilitated the mapping of the end-to-end flow of the work as suggested within DMAIC (Antony, 2006). The team considered bringing in experts in the actual systems, but felt this would lead to the flow being mapped from an "internal" perspective. There was also a concern that the customer focus that had been built up would get lost in the detail, with people wanting to focus on improving what they know, the systems, and ignoring the key focus, the customer.

A workshop was held that simply asked the participants to brainstorm and list the following:

(1) Where do you get your work from, and how?

(2) Where do you send your work to, and how?

From this information the team mapped a process (Figure 7). This was shared and iterated with the team to ensure it was accurate. The team also applied their demand analysis, gathered from the Measure and Analyse steps of DMAIC, to the map. Next, a plan was developed to address the issues highlighted in the Capability of Response (Table IX) showing the issues identified and the actions to resolve them are shown in the plan. The team also measured the volume of work that flowed through their offline queues, what they resolved online and what the passed to the offline teams and how long it took, all related to their purpose statement (from the Define step).

Control. Based on Nonthaleerak and Hendry's (2008) critique of the DMAIC Control step and Table I the team delivered their action plan within one month and the impact

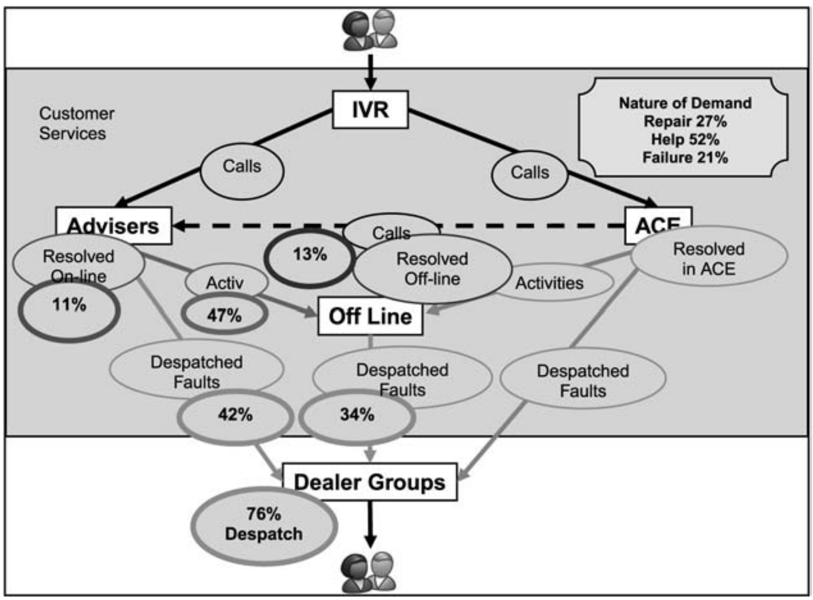

Figure 7.
Call centre 2 repair flow and capability to respond

was tracked against the existing internal measures and the newly devised customer measures.

Looking at three new measures the team used the run charts to track the impact of the action plan delivered against these. First, capability to resolve issues for customers increased from 9.5 to 20.4 per cent when the first changes were made. This has subsequently increased to 26.4 per cent (Figure 8). However, from December 2007 there has been a slight downturn in this performance due to a change in the IT systems used to deal with customer issues.

The percentage of Help work sent to and resolved offline decreased from a mean of 18.7 per cent to mean 6.2 per cent. This meant the repair team no longer needed to deploy resource (2 people) each day to monitor these offline queues and call customers back. The team set up a rota where one person checked the queues for 15 minutes in the morning and again in the afternoon and all employees could be utilised in taking incoming calls.

The volume of activities sent to hold queues to be monitored for call backs reduced from a mean of 1,082 faults to a mean of 338 (Figure 9). Because the team continue to monitor performance this way, they continue to focus on removing waste. The team revisited this process in July 2007, further reducing the volume to a mean of 225. This reduction in volumes being sent offline and the subsequent reduction in the variation of the volumes has been maintained to the last measurement date of March 2008 (Figure 10).

From the analysis it is seen that the process has been improved from the customers' perspective. Customers no longer had to call back if they resolved the fault themselves, they were no longer quoted charges when not applicable, and their faults were detected and directly dispatched by the adviser instead of being placed in a hold queue for further analysis. The variation of the volume and percentage of work being sent to offline queues or despatched also reduced, meaning the repair, engineering and offline teams could better predict their workload, thus enabling the correct number of people to be assigned to these tasks.

In order to test the overall effectiveness of the double DMAIC approach (Table I) existing internally focused measure were also checked. The researcher tracked the impact on existing measures incorporating calls offered, calls abandoned and CHT and PCA15 per cent that management focused upon (Figure 11).

During the improvement period from August to October 2006 the Group was severely damaged by lightning strikes, strong winds and fallen trees. Calls offered increased due to these adverse weather conditions by an average of 1,000 calls per week. In previous instances of storm damage the repair management team implemented a contingency plan to deal with the forecast increase in calls and keep the ABN per cent of calls within target. This involved utilising back office employees or other employees trained in handling simple repair calls. During the proof of concept period (Phase 2) however, this contingency plan was not required, due to the fact that the amount of activities being sent offline had reduced dramatically, enabling the repair team to utilise their two regular offline advisers in online activity.

Within the same period the team were able to reduce their abandon call per cent mean from 5.1 per cent to 4.1 per cent, as highlighted in Figure 12, the best performance for over two years even though call volumes were higher. The team answered these calls without adding additional external resource to their team. Considering that ABN call <5 per cent target is the key measure for call centres and the team achieved a

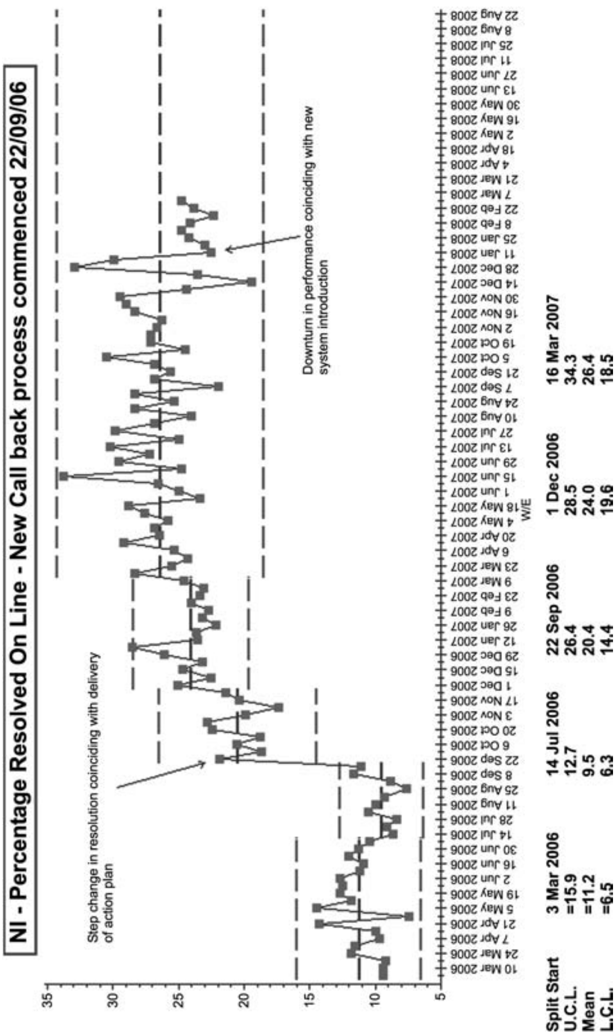

Figure 8.
Repair online resolution tracker

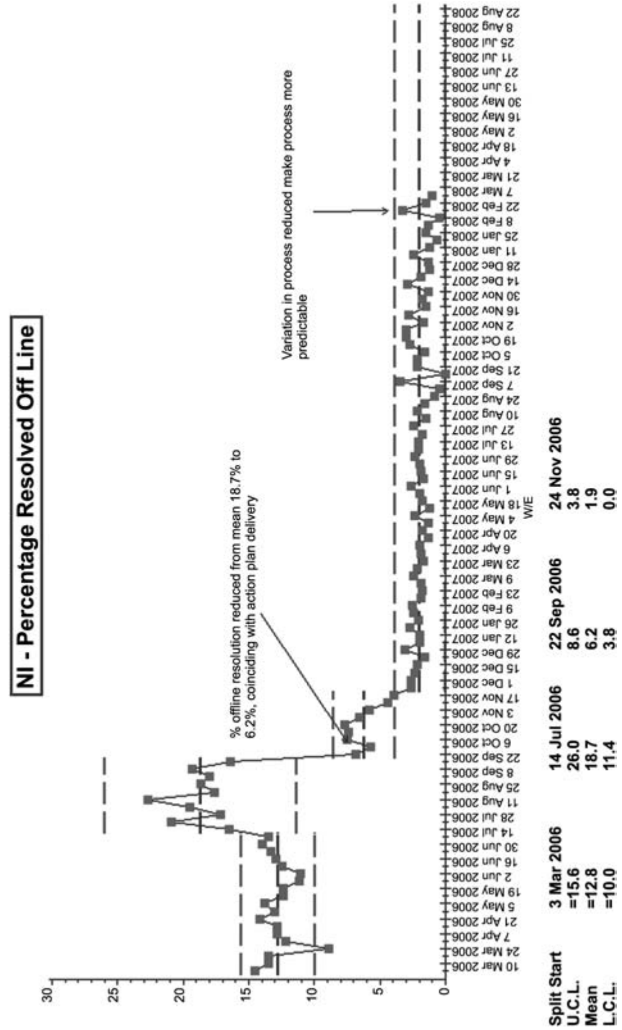

Figure 9.
Repair per cent sent and resolved offline

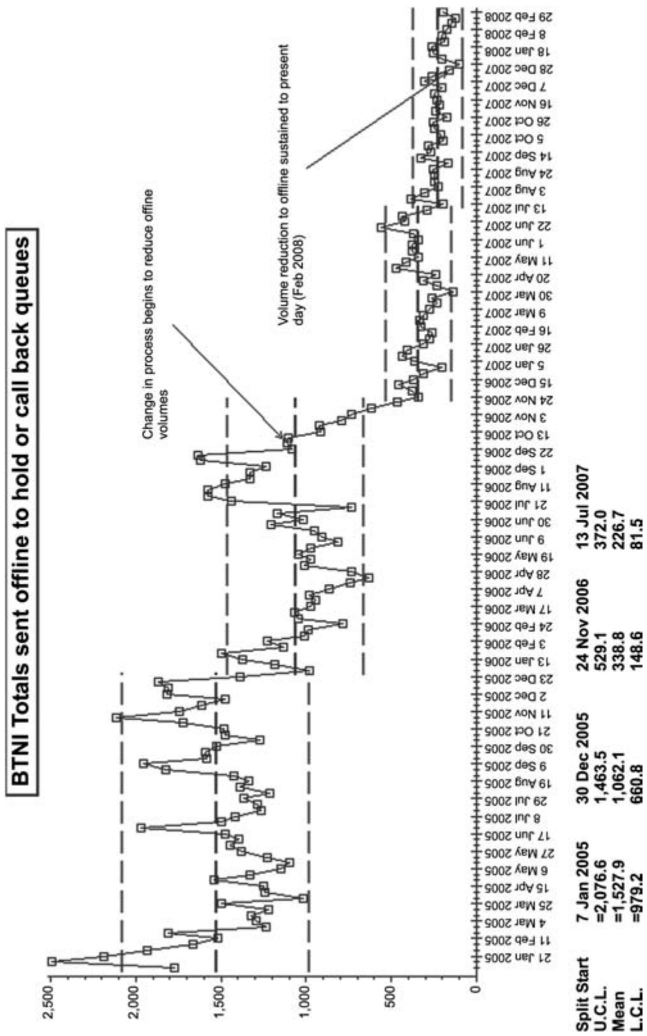

Figure 10.
Repair volumes sent to hold queues awaiting call back to/from customer

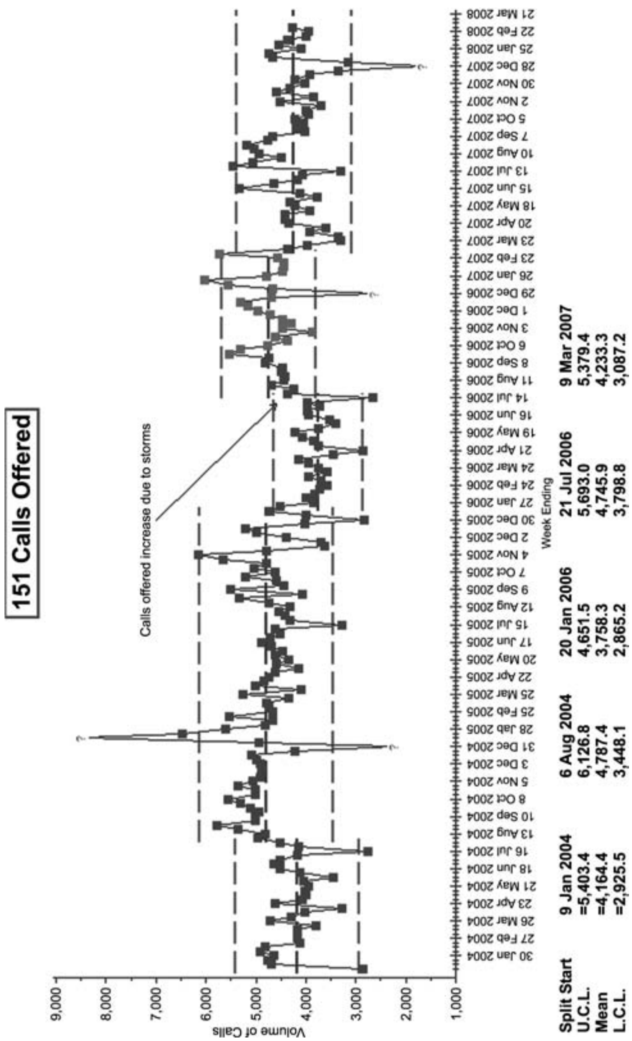

Figure 11.
Repair calls offered from January 2004 to January 2008

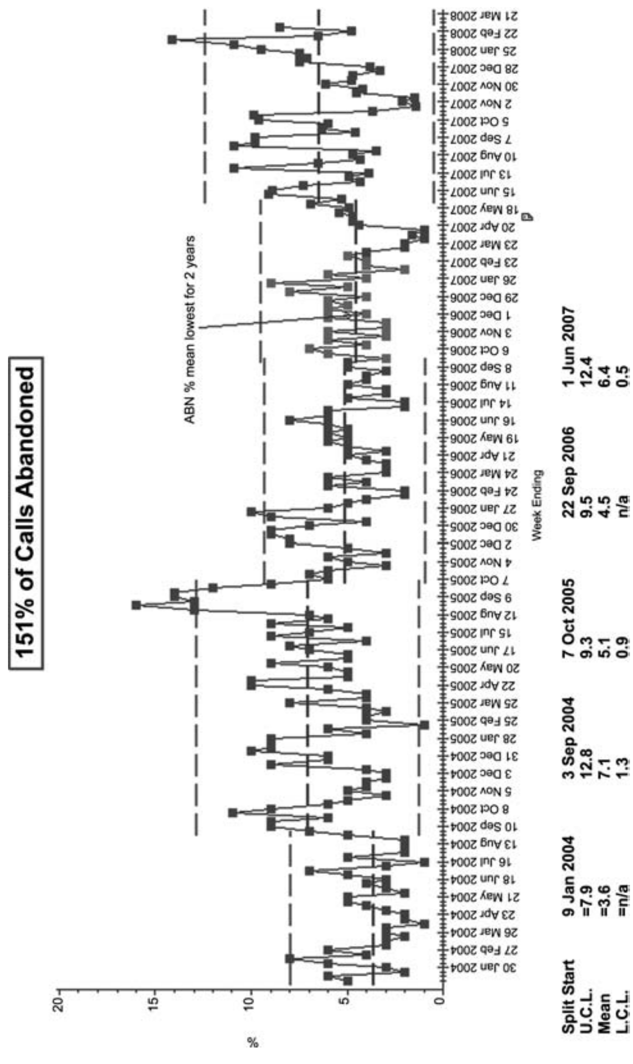

Figure 12.
Repair per cent abandoned calls (ABN per cent)

better than target average during high call volumes and over the winter period, which historically brings higher call volumes ensuring higher ABN rates, the management team determined the proof of concept trial to be a success and agreed to embed the changes within the Group as suggested by Amelsberg (2002) and Nonthaleerak and Hendry (2008).

The reduction of call backs from customers meant that call handling time (CHT) reduced within the same period even though it was envisaged that this would increase as advisors took more time to correctly diagnose the customers issues to ensure it was handled and despatched correctly (Figure 13). Further analysis showed that because the team had reduced the failure demand less customers called back with failure related calls, meaning advisers did not have to spend time explaining, investigating and dealing with the reasons for failure, a process that normally meant putting the customer on hold in order for the adviser to investigate.

As part of the proof of concept the repair advisors also developed a comprehensive training plan linked to the key strategic goals of the organisation and customer expectations similar to that of Taylor *et al.* (2003), ensuring they did not revert back to their previous way of working.

Conclusions and recommendations

From a theoretical perspective a Six Sigma approach has been developed which addresses business improvement issues at a strategic and operational level within organisations, while maintaining the DMAIC methodology. This approach, referred to as the double DMAIC approach to Six Sigma uses systems thinking concepts to ensure that the effects of changes made at a given level are understood at all levels and areas within an organisation.

Applying this theory in phase 1 to call centre 1 has shown that improvement of localised and internal performance measures using operational DMAIC approaches can be misleading and these improvements can be offset by deterioration in performance at other levels in the system or organisation (Curry and Lyon, 2008; Taylor *et al.*, 2003).

In phase 2 the full application of the double DMAIC approach enabled changes to be made and measured from a customer perspective, as suggested by Amelsberg (2002), which in turn had a beneficial effect on internal performance measures. By understanding what matters to customers and identifying and removing the barriers to delivering what matters, call centre 2 removed waste from the system, allowing offline advisors to move to answering calls and resolving customer issues at the first point-of-contact. The results were shared with the repair senior management team as part of the proof of concept review. The review determined that in addition to the improvements highlighted above, the repair team also reduced costs by increasing their capability to resolve issues at the first point-of-contact. Analysis carried out by the resource team supported this finding and concluded that work being handled by the team of 26 advisers during the proof of concept was equivalent to the work of 30 advisors before the proof of concept was carried out, an increased capacity of 14 per cent, equalling reduced costs.

These figures show the need for, first, more research on Six Sigma in service organisations such as call centres (Antony, 2006; Chakrabarty and Tan, 2007) and second, the need for further applications and refinement of the double DMAIC approach at strategic and operational levels (Friday-Stroud and Sutterfield, 2007).

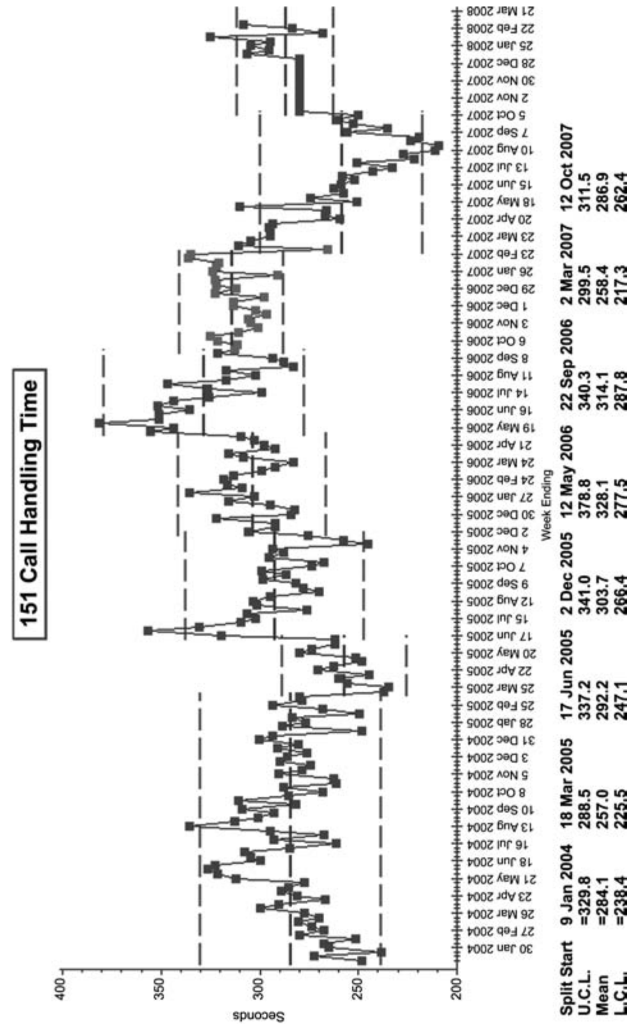

Figure 13.
Repair call handling time (CHT) performance

References

Amelsberg, J. (2002), "Systematic performance and cost management: a management framework for organizational excellence", *Proceedings of the Annual Quality Congress, Milwaukee, WI*, pp. 487-500.

Antony, J. (2006), "Six Sigma for service processes", *Business Process Management Journal*, Vol. 12 No. 2, pp. 234-49.

Antony, J., Antony, F., Kumar, M. and Cho, B. (2008), "Six Sigma in service organisations", *International Journal of Quality & Reliability Management*, Vol. 24 No. 3, pp. 294-311.

Bellman, B. (2007), "Tools for first call resolution – making contact centres more efficient and responsive", *Business Communications Review*, Vol. 37 No. 12, pp. 28-31.

Chakrabarty, A. and Tan, K. (2007), "The current state of Six Sigma application in services", *Managing Service Quality*, Vol. 17 No. 2, pp. 194-208.

Cleveland, B. (2007), "The measures every successful call centre should have – there are seven types of measures that should be in place in every customer contact centre", *Call Center Magazine*, Vol. 20 No. 4, pp. 6-10.

Contact Babel (2006), *The UK Contact Centre Operational Review*, 4th ed, Contact Babel, Brighton.

Conz, N. (2007), "How can we help you? Insurers are changing their contact center philosophies in an effort to improve customer satisfaction and boost customer retention", *Insurance & Technology*, Vol. 32 No. 9, p. 35.

Curry, A. and Lyon, W. (2008), "Call centre service quality for the public: a Scottish framework for the future", *Managing Service Quality*, Vol. 18 No. 2, pp. 194-208.

de Koning, H. and de Mast, J. (2006), "A rational reconstruction of Six Sigma's breakthrough cookbook", *International Journal of Quality & Reliability Management*, Vol. 23 No. 7, pp. 766-87.

Eisenhardt, K.M. (1989), "Building theories from case study research", *Academy of Management Review*, Vol. 14 No. 4, pp. 532-50.

Friday-Stroud, S. and Sutterfield, J. (2007), "A conceptual framework for integrating Six Sigma and strategic management methodologies to quantify decision making", *The TQM Magazine*, Vol. 19 No. 6, pp. 561-71.

Gettys, R. (2007), "Using lean Six Sigma to improve call centre operations", iSixSigma.com, available at: http://finance.isixsigma.com/library/content/c070418a.asp

Harney, B. and Jordan, C. (2008), "Unlocking the black box: line managers and HRM-performance in a call centre context", *International Journal of Productivity and Performance Management*, Vol. 57 No. 4, pp. 275-92.

Hart, M., Fichtner, B., Fjalestad, E. and Langley, S. (2005), "Contact centre performance: in pursuit of first call resolution", *Management Dynamics*, Vol. 15 No. 4, pp. 17-29.

Koh, S., Gunasekaran, A., Thomas, A. and Arunachalam, T. (2005), "The application of knowledge management in call centres", *Journal of Knowledge Management*, Vol. 9 No. 4, pp. 56-70.

McAdam, R. and Lafferty, B. (2004), "A critique of Six Sigma: statistical measures or strategic change?", *International Journal of Operations & Production Management*, Vol. 24, pp. 530-49.

Morgan, S. and Cooper, C. (2004), "Shoulder work intensity with Six Sigma", *Nursing Management*, Vol. 35 No. 3, pp. 28-33.

Nonthaleerak, P. and Hendry, L. (2008), "Exploring the Six Sigma phenomenon using multiple case study evidence", *International Journal of Operations & Production Management*, Vol. 28 No. 3, pp. 279-303.

Peltokorpi, A. and Kujala, J. (2006), "Time-based analysis of total cost of patient episodes: a case study of hip replacement", *International Journal of Health Care Management*, Vol. 19 No. 2, pp. 136-45.

Proudlove, N., Moxham, C. and Boaden, R. (2008), "Lessons for lean in healthcare from using Six Sigma in the NHS", *Public Money and Management*, February, pp. 27-34.

Remenyi, D., Williams, B., Money, A. and Swartz, E. (1998), *Doing Research in Business and Management: An Introduction to Process and Method*, Sage, London.

Revere, L. and Black, K. (2003), "Integrating Six Sigma with total quality management: a case example for measuring medication errors", *Journal of Healthcare Management*, Vol. 48, pp. 377-85.

Seddon, J. (2001), *The Vanguard Guide to Transforming Call Centre Operations*, Vanguard Education, London.

Seddon, J. (2005), *Freedom from Command and Control*, Vanguard Education, London.

Sehwail, L. and DeYong, C. (2003), "Six Sigma in health care", *International Journal of Health Care Quality Assurance*, Vol. 16 No. 4, pp. 1-6.

Talyor, P., Baldry, C., Bain, P. and Ellis, V. (2003), "A unique working environment: health, sickness and absence management in UK call centres", *Work, Employment and Society*, Vol. 17 No. 3, pp. 435-58.

Yin, R.K. (2003), *Case Study Research: Design and Methods*, Sage, London.

Further reading

Dahlgaard, J. and Dahlgaard-Park, S. (2006), "Lean production, Six Sigma quality, TQM and company culture", *The TQM Magazine*, Vol. 18 No. 3, pp. 263-81.

Corresponding author

Rodney McAdam can be contacted at: r.mcadam@ulster.ac.uk

Received 7 October 2008
Revised 14 January 2009
Accepted 20 January 2009

Continuous improvement: role of organisational learning mechanisms

Judy Oliver
Swinburne University, Melbourne, Australia

Abstract

Purpose – The purpose of this study is to explore the use of the performance measurement system as an organizational learning mechanism to support continuous improvement.

Design/methodology/approach – The paper reports the results of a survey of Australian organizations certified to quality standard ISO 9000.

Findings – For those respondents who consider their organization's quality program to be successful, the findings indicate that such organizations have embedded quality into the culture of the organization, and have developed performance measurement systems as an organizational learning mechanism to support the continuous improvement initiatives.

Practical implications – The paper highlights the need for management to ensure that the organization's management control systems are structured to support continuous improvement initiatives.

Originality/value – The paper explores the links between continuous improvement and organizational learning.

Keywords Continuous improvement, Workplace training, Performance measures, Quality management, Customer satisfaction

Paper type Research paper

Introduction

Quality management has become recognised as one of the key strategies for organizations to improve their productivity and international competitiveness and thereby meet the demands of customers (Chenall, 1997; Czuchry *et al.*, 1997; Evans and Lindsay, 1996; Lee and Walden, 1998; Sambrook and Stewart, 2000; Spong, 1994). However, the literature reports mixed success of quality management programs and it is suggested that 60 percent to 80 percent of attempted implementations failed to meet their objectives (Lau and Anderson, 1998, p. 85). This high level of reported failure would be of concern to stakeholders given the considerable amount of resources that are invested in quality initiatives (Srinidhi, 1998). One of the biggest causes of failed quality initiatives has been a lack of management and employee commitment. Perhaps the greatest cost for an organization as a result of failure is the loss of morale or an increase in cynicism among employees (Dooley and O'Sullivan, 1999). Lau and Anderson (1998) suggest that for a quality program to be successful an organization needs to have: a strategy that guides the quality initiative; an infrastructure, such as a performance measurement and evaluation system, to monitor and control quality programs; and the necessary cues to encourage organizational learning.

The purpose of this paper is to report the results of a survey of managers of Australian organizations certified to quality standard ISO9000. The survey was

International Journal of Quality & Reliability Management
Vol. 26 No. 6, 2009
pp. 546-563

0265-671X
DOI 10.1108/02656710910966129

structured to seek the perceptions of managers on a range of issues in relation to their organization's quality initiatives. In particular, the focus was on the investigation of the role of organizational learning to support continuous improvement, and to identify the organizational attributes, in particular organizational learning mechanisms, for those organizations where management view the quality program as successful. The mixed success of quality programs as reported in the literature provides the motivation for this study.

The paper is organized as follows: The next section provides a discussion of the quality management and organizational learning and associated links. This is followed by a discussion of the role of the performance measurement system to support continuous improvement. In the next section an outline of the research method is detailed. Then the survey results are discussed and are then followed by some concluding comments.

Literature review

Quality and learning. There is no universally accepted definition of quality and, as such an organization will need to develop its own working definition that will find its origin in the organization's vision (Chapman *et al.*, 1997; Groth, 1995; Lau and Anderson, 1998; Sinclair and Zairi, 1995; Reeves and Bednar, 1994; Srinidhi, 1998). Therefore, the meaning of quality will be peculiar to individual organizations with different definitions of quality appropriate under different circumstances (Reeves and Bednar, 1994). During the evolution of quality the terminology used to describe the quality movement changed without any clear declaration and at some point the term total quality management (TQM) began to be used instead of total quality control or just quality control (Dahlgaard, 1999). Today TQM is the term generally used to describe quality practices within organizations. TQM can be regarded as a management approach characterised by three core principles: customer focus; continuous improvement; and employee involvement (Dean and Bowen, 1994; Evans and Lindsay, 1996; Sitkin *et al.*, 1994; Yong and Wilkinson, 2001).

It is suggested that the success of TQM is dependent on an organization's ability to learn, to absorb, to adapt and to apply conceptual changes and integrate them throughout the organization (Ford, 1991, cited by Terziovski *et al.*, 2000). The ability to learn new sets of skills on a continuing basis represents a sustainable source of advantage for the future (Liedtka and Rosenblum, 1996; Sambrook and Stewart, 2000; Tranfield *et al.*, 2000); suggesting continuous improvement will be achieved if learning takes place within the organization (Bessant and Francis, 1999; Egan, 1993). A learning focus will encourage employees to provide feedback to evaluate performance, enabling the outcomes of the continuous improvement activities to be incorporated into the knowledge base within the organization. From this knowledge base, it allows future improvement to be built on past accomplishments (Jha *et al.*, 1996).

Organizational learning has been described as the process of improving actions through better knowledge and understanding (Fiol and Lyles, 1985; Huber, 1991). With the objective of maintaining or improving performance based on experience (Wick and León, 1995; DiBella and Nevis, 1998). It can be viewed as a characteristic of an organization that is observed through the actions of the parts and describes certain types of activities or processes that may occur at several levels of analysis (for example, individuals, teams and companies). Organizational learning can be found in any organization, but the learning organization will embody organizational learning in all its actions and exemplifies the ideal application of organizational learning. Learning organizations are those that purposefully construct structures and strategies so as to enhance and maximise organizational learning (Dodgson, 1993, p. 377).

Terziovski *et al.* (2000) carried out field research to examine the mutual dependence between TQM and the learning organization. They concluded that the success of the companies' quality programs was due to the sustained commitment to "learning" and will involve the process of building procedural knowledge, cognitive strategies and attitudes. Learning can concentrate on methods and tools to improve what is already being done, known as single-loop learning, or on testing the assumptions underlying what is being done, known as double-loop learning. Organizations may have a preference for one mode over the other, but a sound learning system requires both approaches (Appelbaum and Reichart, 1997).

Organizational learning is operationalised through organizational learning mechanisms (OLMs), which are the institutionalised structural and procedural arrangements that aid the learning process (Lipshitz and Oz, 1996). Such mechanisms allow organizations to collect, analyse, store, disseminate and use information that is relevant to the organization. It is due to the existence of such "mechanisms" that organizational learning can be studied as an actual phenomenon. OLMs enable the experiences of individual organizational members to be analysed and shared by other organizational members. The experience becomes the property of the entire organization through distribution of lessons learned to relevant units or through changes in standard operating procedures (Lipshitz and Popper, 2000).

Therefore, learning can be viewed as the foundation for improvement activities. It provides the organization with the capabilities to take action and without which any attempts at improvement will possibly fail (Bessant and Francis, 1999; Wick and León, 1995). Learning can assist an organization in its quest for continuous improvement by helping to avoid repeating mistakes; building sensitivity to the changing world so that the organization can adapt better; and improving operations by understanding the weaknesses in the past and identifying how to correct them (Lee, 1995). Learning will be seen to have occurred when an organization performs in changed and better ways (Dodgson, 1993). Perhaps the underlying reason behind the lack of success of some quality programs is that the processes put in place lack the necessary cues for quality learning. This research may contribute to an understanding of whether a lack of learning is inhibiting the success of quality programs.

Management control system. The management control system (MCS) can be viewed as the organization's control package, with components such as of the accounting information system (cost systems and budget systems), performance measurement and reward systems and planning systems. However, in reality it could be any system to monitor and assist work practices. An organization's MCS can empower organizational learning through its design features and interactively influence strategy (Simons, 1990) and thereby act as an OLM. Control is the continuing process of evaluating performance and taking corrective action when necessary and enabling the organization to maintain high quality processes, and also to bring processes under control in order for improvements to be made (Evans and Lindsay, 1996). As noted by

Simons (1991, p. 49) the MCS is influential on organizational activity as it represents "…the formalized routines and procedures that use information to maintain or alter patterns in organizational activity …". Therefore, if the MCS is structured to support the learning environment it should be a system that supports decision-making, facilitates rapid and effective learning and unlearning, and enables the acquisition and development of information, knowledge and understanding.

The MCS sets a framework for an organization's information seeking, accountability and feedback designed to ensure that it adapts to changes in its environment (Kloot, 1997; Lowe, 1971). As noted by Simons (1991) control systems allow employees to access information to undertake their tasks, and also provide direction in the accomplishment of those tasks by providing information necessary for feedback and control. MCS are an important element of strategy implementation by translating the plans into action (Simons, 1992). The MCS should support and put into practice the operating philosophies of continuous improvement, and be adaptable to revision whenever changes are made to the operating strategy (Banker *et al.*, 1993; Bessant and Francis, 1999).

The ability of an organization to adapt successfully to changes in the competitive environment can be seriously inhibited by a poorly designed performance measurement system (Sinclair and Zairi, 1995). "Ownership" of the measurement system is important to embed the behaviour that promotes improvement (Bessant and Francis, 1999), and those directly involved in the continuous improvement process should be involved in its operation and implementation. The performance measurement system is a key enabler to encourage improvement as it gives focus to improvement activities and assists in the identifying the extent to which performance has changed (Bessant and Francis, 1999). This can be achieved if an organization is able to define, in specific performance terms, what it means by quality and then to measure these performance variables objectively (Krishnan *et al.*, 1993). Without an appropriate performance measurement system, improvement activities can fail (Banker *et al.*, 1993; Chapman and Hyland, 2000). It is suggested by Oakland (1993) cited by Sinclair and Zairi (1995) that appropriate performance measurement systems play the following roles in quality and productivity improvement: ensures customer requirements have been met; provides standards for establishing comparisons; provides visibility and provides a "scoreboard" for people to monitor their own performance levels; gives an indication of the costs of poor quality; justifies the use of resources; and provides feedback for driving the improvement effort. Employees should be able to monitor and (if necessary) change their actions based on the feedback gained from comparing actual performance against target. For example, Chapman and Hyland (2000) concluded from a study of small-to-medium Australian manufacturing organizations that there was a low level of correlation between the competitive measures and the motivation for continuous improvement or content of the continuous improvement program. They also identified that the measurement system often did not include a closed feedback loop and any learning that did take place was usually localised owing to the absence of any effective information collection and storage mechanisms.

Therefore, a well-structured measurement system provides the linkage between strategies and actions. The links are established by the performance goals developed to encourage employee behaviour to meet the organization's objectives and facilitate and support induced quality learning by incorporating goal-setting feedback as an essential component of the system. Goals are broad statements that set the direction for the organization in realising its mission and closing the gap between where it is and where it wants to be (Evans and Lindsay, 1996). The goals need to be consistent with the key factors that drive the business and must not undermine quality. Fine (1986) argues that to achieve cost reduction and productivity improvement the performance measurement system should support quality-based learning by making use of frequently revised goals. Further support to the benefits of performance goals is found in the goal setting literature which identifies that individuals with specific and hard or challenging goals outperform individuals with specific easy goals, do-best goals, or no assigned goals (Dossett *et al.*, 1979; Locke *et al.*, 1981). A review of all available experimental field studies on goal setting found that when goals are set the median improvement in productivity and quality was 16 percent and when combined with monetary incentives, median performance was improved by more than 40 percent (Locke *et al.*, 1981).

Quality goals are the central focus of an effective quality program and should be supported by a strong measurement system and must be quantifiable (Lau and Anderson, 1998). Such quantitative measures allow specific goals to be established and specific results to be forecast and also provide the basis for clear company-wide quality discussions at all levels of the organization. This will provide a higher level of precision for discussing results. If such measures are clear it should lead to worker acceptance and commitment. To provide more meaning to employees, the organization needs to translate the quality goals into operational goals.

Chapman *et al.* (1997, p. 433) identified best practice in relation to performance measurement by examining organizations that have won Australian quality awards (Chapman *et al.*, 1997, p. 433). Best practice attributes identified included:

- Goals, priorities and targets, which are clear and unambiguous to all employees. These have been deployed throughout the organization while retaining alignment to organization-wide improvement strategies.
- Quantifiable goals with measurement/benchmarking processes to provide clear indications of progress towards the goal.
- Competitor benchmarking in the area of customer satisfaction is a continuing activity and the information is fed into the strategy and goal setting process.
- Data collection and analysis relating to key internal processes are a fundamental part of routine work. Results of such measurement are used to produce revised goals and targets.

Specific performance goals or targets to support quality have also broadened with the move from quality control to TQM (Dahlgaard, 1999). Initially, quality was measured in defect rates, complaint rates, returns, etc., and has now extended to measures with a focus on customers and employees. However, it has been suggested that the poor performance of many new TQM initiatives can be accounted for by the continued reliance on out-dated traditional performance measures that focus on the financials. As mentioned by Oakland (1993) the key success factors today are not easily found in the financials, and the focus should be on customer satisfaction and non-financial information relating to the work effort and to costs relating specifically to quality. It is suggested that goal conflicts can be avoided by ensuring that goals are consistent, subsume other goals and are sequentially prioritised. Consensus on what goals to pursue helps to avoid confusion caused by simultaneously pursuing multiple quality programs (Krishnan *et al.*, 1993). The goals need to be consistent with the key factors that drive the business and must not undermine quality. For example, Lincoln Electric, a US manufacturer, gave employees no credit for units that did not meet the quality standard so as to ensure there was no quantity/quality trade off (Wright, 1994). Daniel and Reitsperger (1992) suggest if quality is a strategic priority then the provision of quality targets and feedback to operating management should reflect the importance of quality improvement and emphasise the importance of continuous improvement.

A study undertaken in New Zealand explored the changes in the management accounting system (MAS) in relation to performance measures as a result of a TQM implementation (Hoque and Alam, 1999). Pre-TQM the organization's MAS was historical and financial accounting orientated and post-TQM the organization recorded both financial and non-financial events of the company. Managers from the research site "expressed a high degree of satisfaction with the post-TQM MAS (management accounting system)….. helped them coordinate, plan and communicate the TQM related work to the best interests of the company".

The research questions

Deming (1993) proposed a "theory of profound knowledge" and a key aspect of this theory is that the success of quality management efforts depends on the effective integration of various management sub-systems (Waldman, 1994, p. 33). Therefore, the maximum effectiveness of TQM may be dependent on whether the performance management sub-systems are consistent and integrated with continuous improvement sub-systems. Organization's seeking advancement must have a set of metrics to quantify both the efficiency and effectiveness of actions.

It is argued that for a continuous improvement philosophy to be successful the organization must have in place both the commitment to learning and the control systems that are flexible enough to meet the changing needs of the business environment. The absence of such attributes may account for the lack of success of some quality programs. It is important for management to recognise the need to create the environment, which will encourage learning to achieve continuous improvement, and to have the MCS act as an organizational learning mechanism. As noted by Berling (2000) "the task is not only to start the improvement process, but also to sustain it and to incorporate it into the normal part of everyday work".

The decision to apply a continuous improvement philosophy to all activities within the organization in itself implies a learning approach, as the organization would be focused on improving the way it currently operates its business. However, the learning must be nurtured and encouraged by the practices adopted within the organization. OLMs will be instrumental in developing the learning culture, and the form and extent of the adoption of OLMs will be a determinant of the level of success achieved by an organization in its continuous improvement endeavours. Therefore, the research questions emanating from this general proposition are:

RQ1. What is the motivation for an organization to adopt a quality approach to its operations?

RQ2. What are the characteristics of the organizational learning mechanisms used by quality-focused organizations to support continuous improvement?

RQ3. What are the characteristics of the organizational learning mechanisms favoured by organizations with a more successful quality program?

Methodology

The empirical research will focus on organizations that have adopted a quality focus as it would be expected that such organizations are more likely to have an operating philosophy of continuous improvement (Abraham *et al.*, 1997; Terziovski *et al.*, 2000). Organizations that have ISO 9000 certification were selected as such certification provides an independent third-party assessment that the organization has implemented a quality approach to its operations, at least in terms of the requirements of the quality standard; and the use of such organizations is seen in the research of others (Claver *et al.*, 2002; Hendricks and Singhal, 1997; Llopis and Tari, 2001). Quality managers and finance managers were considered to be the most appropriate respondents as their work responsibilities would expose them to their organization's operating practices in relation to quality management and the use of control systems.

Potential respondents were sourced from the publicly available on-line Joint Accreditation System of Australia and New Zealand (JAS-ANZ) database of certified organizations. A random sample, from each state of Australia, was selected comprising 500 organizations.

Surveys were posted to both the Quality Manager and Finance Manager of each organization representing a mail out to 1,000 managers, and a follow-up mail out was undertaken two months later. A total of 16 questionnaires were received marked "return to sender" which left the potential for nine hundred and eighty four managers to respond. In all, three hundred and three responses were received representing a 30.6 percent response rate, and of these two hundred and seventy seven represented usable responses, resulting in a usable response rate of 28.2 percent. A usable survey was deemed to be one in which the respondent answered all but a few questions. To test for non-response bias an independent sample *t*-test was conducted comparing early and late respondents. The null hypothesis was posed that the samples came from the same population and for all characteristics, expect gender; the null hypothesis is not rejected at the 0.05 percent level of confidence. Further examination of this result shows that in the second mail out fewer males responded. However, this was not considered significant because of the high percentage of female responses overall.

Discussion

Profile of respondents

The majority of respondents (62.6 percent) is responsible for quality management activities, with 26.8 percent responsible for financial management and 10 percent responsible for general management, with a further two surveys completed jointly by the Quality Manager and the Finance Manager. The majority of respondents are female

(80.9 percent), of whom 138 are responsible for quality management; 62 for financial management; and 20 have general management responsibilities. Of the males responding (19.1 percent), 32 are responsible for quality management; 11 for financial management; and 8 have general management responsibilities. The majority of respondents (84.1 percent) have more than 10 years experience in business with 77.1 percent of respondents having undertaken post-secondary education. Such a profile suggests that respondents have the working knowledge to comment on the issues raised in the questionnaire in relation to their organization and its quality initiatives.

Profile of organizations

The number of employees is used as a measure to determine the size of respondents' organizations, with the findings showing 49.6 percent small- to medium-sized organizations (under 100 employees); 31.5 percent medium sized organizations (101 – 500 employees); and 18.9 percent large organizations (over 500 employees). The majority of the respondents (97.7 percent) consider their organizations are operating in a competitive environment, with 59.2 percent of the respondents rating the environment as very competitive. The majority of respondents (57.7 percent) view their organization's product/service as superior quality to competitors, with 41.2 percent ranking it as similar and only one respondent giving a ranking below competitors. Two respondents from Government organizations did not rank their organization.

The majority of respondents (96.3 percent) identify quality-related factors as the main source of competitive advantage. In particular, the nominated sources of competitive advantage by respondents are: flexibility in responding to customer needs (24.2 percent), higher quality (of product/service) than competitors (23.0 percent), and product/service differentiation (18.3 percent). A combination of these factors was nominated by 24.0 percent of respondents, with another 10.5 percent identifying a combination of both quality and cost related factors. As respondents are from organizations with ISO 9000 certification it would be anticipated that a quality approach would influence the development of competitive strategies.

Quality

For the majority of respondents' organizations (54.5 percent), quality has been important for more than 10 years, with a further 35.5 percent stating that quality has been important for more than five years. Quality has been adopted organization-wide for the majority of respondents' organizations (86.8 percent) with the remaining (13.2 percent) noting its influence at either the individual work-unit level or project level. In relation to the outcomes of the quality program the majority of respondents (82.76 percent) consider overall performance has improved, with 75.78 percent noting an improvement in their organizations' competitive position. In fact, 80 percent of respondents indicate that their organization would not have performed as well without a quality program.

The majority of respondents (86.1 percent) agree that top management in their organization is committed to the quality program; and 83.4 percent of respondents note that management ensure employees are aware of what quality means to the organization. However, only 53.5 percent of respondents consider that management view quality as the way to increase profits. To further explore this respondents were asked to comment on whether management viewed results as more important than processes. The responses show that 22.8 percent of respondents agree with the statement, 45.5 percent disagree and 31.6 percent are uncertain. This could raise issues about management's full commitment to the quality program. The ownership of the quality program is not clear, as only 58.7 percent of respondents indicate that employees believe that quality is their responsibility. Given that employees are an important key to success this bare majority is a concern as this relatively low level of ownership could inhibit the diffusion of quality management within the organization. To be successful quality programs require the support of both management and employees. Selected comments from respondents that further explain these findings are:

> [. . .] (lack of) support of supervisors and middle management . . .

> [. . .] moderately high standard of quality but needs to be pushed to get the people in the field to keep it ongoing. They think of it as a burden (cost) not a tool to help . . .

> [. . .] the systems (as advised by external parties) are at the cutting edge. While senior management support is excellent intentionally it is not necessarily practised and driven with the same support. Middle managers give moderate to nil support . . .

Learning environment

It is important for the organization to create the environment that will encourage individuals to learn, which, in turn, may lead to organizational learning. As noted by Ahmed *et al.* (1999) a learning and continuously improving company requires an organizational culture to guide employees. An understanding of an organization's learning orientation, that is, its values and practices, will identify whether the environment created will encourage learning. To build a shared vision within the organization, that is, to give everyone a common identity and sense of destiny, it is important for management to provide direction. This direction can be given by the performance goals to guide activities, the organization's values and its mission statement.

Respondents were asked to comment on particular values (espoused theories) in their organization. The majority of respondents (93.5 percent) identify continuous improvement as an important goal for their organization and this is reinforced by 86.5 percent of respondents noting its influence when formulating the strategic plan. A goal of continuous improvement implies the need for learning, and 87.0 percent of respondents agree that continuous learning is valued in their organization. These findings suggest that such organizational values should encourage a learning environment in respondent organizations. However, are these values supported within the organization?

The majority of respondents (51.7 percent) agree that "what gets said gets done", that is, the espoused theories are in line with the theories in action. However, the findings also indicate that 33.5 percent of respondents are unsure, which could suggest that management inactions are inhibiting learning. A bare majority of respondents (50.2 percent) disagrees that operational planning only involves managers, which questions the participation of employees in the planning process in some respondent organizations. The encouragement for continuous improvement is noted by 50.9

percent of respondents who agree that their organization encourages employees to look at other approaches to organizational activities. But, once again, the findings show that 30.7 percent of respondents are unsure about management's attitude.

Learning can be seen to occur when organizations perform in changed and better ways. The majority of respondents (96.4 percent) identify the importance of their organization being adaptable to changes in the business environment. Learning will involve the process of building procedural knowledge, cognitive strategies and attitude. The majority of respondents (85 percent) agree that employees are encouraged to work smarter not harder, which is in line with the motivation for management to improve performance. This suggestion is reinforced by 84.5 percent of respondents who agree that their organization encourages employees to question current work practices and find new ways of doing things. In addition, 80.2 percent of respondents agree that in their organization standard operating procedures are reviewed regularly, which would encourage improvement in existing capabilities and promote single-loop learning and also may encourage double-loop learning from the review process. 74.2 percent of respondents note that managers do not punish mistakes but encourage employees to explore alternatives. Further encouragement is given to employees as 70.7 percent of respondents agree that employees have been given decision-making responsibilities to deal with problems relating to specific work activities.

However, despite the encouragement only 58.7 percent of respondents identify that employees are focused on improving existing capabilities. The empowerment of employees to initiate change is unclear as only 42.9 percent of respondents agree that this is encouraged in their organization, and a further 35.0 percent of respondents are unsure about their organization.

Performance measurement

Respondents were asked to rate the importance of certain factors in the development of the key performance indicators (KPIs) for their organization. Factors ranked as extremely important are customer satisfaction (mean 4.76), process improvement (by way of cost efficiency) (mean 4.38), and profit (mean 4.30). These are in line with the key factors in relation to the motivation for the quality initiatives and indicate that the KPIs are supporting the organizations' objectives. Revenue growth (mean 4.00) and return on assets (mean 3.9) are ranked lower than other factors which may be due to the organizations focusing on maintaining the existing customer base by improving both customer satisfaction and internal processes. Least important is the market share (mean 3.76), which might be explained by many organizations being in a competitive environment thereby reducing opportunities for expansion. Retention of existing customers may be more critical.

For the majority of respondents (89.2 percent) the key performance indicators (KPIs) are an output of the strategic planning process. Respondents' comments noting the importance of developing KPIs are:

> [...] there is a high development process occurring at this time in regards to strategic management (KPIs etc) ...

> [...] KPIs and benchmarking are the two most important areas to ensure continual improvement and financial stability ...

> [. . .] our basic quality system is solid and reliable to ensure quality of product. Our next step is to become more proactive and create more KPIs and reporting on and gauging progress with clearly identified benchmarks . . .

The majority of respondents (81.5 percent) agree that their organization is able to link the operational performance measures to the strategic plan which suggests that employees' effort should be in line with the performance outcomes embodied within the organization's overall strategy. However, 34.3 percent of respondents identify that their organization has difficulty in translating the quality goals into operational goals, which could compromise the success of the quality program in such organizations.

Of the respondents, 80 percent consider that all appropriate management and employees in their organization are made aware of the performance measures to encourage continuous improvement; and 73.3 percent agree that employee involvement in goal setting is important. It is assumed that this is in relation to operational goals as only 20.9 percent of respondents agree that employees are involved in the development of the strategic plan. One comment made by a respondent highlights the importance of this involvement:

> [. . .] more employee involvement in goal setting required – not enough performance review at divisional levels. Require more time to be spent on review of performance on projects and identify lessons learnt and implement correction action and improve systems, procedures etc. Goals should also be communicated to all employees not management and this also applies to performance results – this would increase a sense of ownership and commitment at all levels . . .

Characteristics of performance goals

Performance goals should provide the signals to employees of what actions are required to meet organizational objectives. Respondents (73.3 percent) agree that the performance goals are clear and consistent; with 65.2 percent of respondents suggesting that the performance goals reflect the importance of quality improvement activities, which is consistent with 68.2 percent of respondents agreeing that performance measures encourage employees to work towards quality goals. Of concern is that only 55.7 percent of respondents agree that their organization frequently revise the performance measures to adapt to changes in operating conditions.

It could be expected that the performance goals would encourage the learning style, that is, either double-loop or single-loop learning. Only 49 percent of respondents consider that the performance goals are structured to encourage employees to explore new ways of doing their jobs. The use of stretch goals as a means to encourage improved performance is only noted by 41.6 percent of respondents and only 48.9 percent of respondents agree that performance goals are used to modify employee behaviour. The minority of respondents (44.5 percent) agrees that performance goals promote dialogue and debate among employees with 56.5 percent of respondents who agree that performance goals encourage cooperation and interaction between employees. The above findings raise doubts as to whether the performance goals would encourage double-loop learning.

Planning and control are strengthened by the ability of employees to respond to problems. The evaluation of actual performance against targets is important for the control of operations. The majority of respondents (84.8 percent) agree that performance feedback is important for investigating problem areas. 60 percent of respondents agree that feedback gained from assessing performance against target enables the instigation of rapid corrective action. The ability of employees to react to feedback may be limited as only 60 percent of respondents agree that employees receive regular appraisal and feedback about their work performance. The responses also indicate that despite feedback being important, 28.4 percent of respondents are unsure whether feedback enables rapid corrective action, and 26 percent of respondents are unsure whether employees receive regular feedback.

Factors influencing success of quality initiative

The success of the quality program in respondents' organizations has been mixed. For the majority of respondents (77.4 percent), the quality program has met the expectations set, with 16.1 percent of respondents rating their organization's achievement as exceeding expectations. Of the respondents, 14 (5.1 percent) perceive their organization's disappointment with the outcome of the quality program with a further four respondents unable to determine the outcome of the quality program at the time of the study. It should be noted that the interpretation of success in this study is based on a subjective assessment, given that it represents each respondent's view of success in relation to his or her own organization. It is argued that this is an appropriate approach given that, in the end, if management is dissatisfied with the outcomes of practices/initiatives put into place, then such practices will not have longevity within the organization, despite any external commentator suggesting differently.

Comments from respondents who rate the quality initiative as "fell short of expectations" reinforce the barriers to success identified by Kaye and Dyason (1995), Krishnan *et al.* (1993), Lorente *et al.* (1999), Sitkin *et al.* (1994) and Lau and Anderson (1998). Selected comments highlighting the reasons behind the lack of success of the quality program follow:

> [. . .] we are constantly trying to improve our quality system, but middle management and supervisors still see it as a hindrance not a help. Unwilling to change . . .

> [. . .] poor management understanding and support . . .

> [. . .] lack of resources. . . people, plant and equipment, to maintain quality focus; productive focus instead of a market (customer) focus . . .

> [. . .] we plan well but action/review poorly. . .lack of resources (constant drive to reduce costs by reducing people) leaves little opportunity to explore in-depth continuous improvement . . .

> [. . .] our quality initiatives have been production related rather than encompassing all business activities . . .

> [. . .] conflict in measuring KPI's (quality vs reject rate, on-time delivery); not clearly walking the talk . . .

> [. . .] large organization; quality means different things to different areas. Quality is seen as an input to improve efficiency, not so much an output . . .

Further analysis of the results were undertaken to identify significant relationships between the perception of respondents to a range of factors in relation to their organization and the success of the quality program. An analysis of the most significant variables suggests that there is a relationship between the level of success of the quality program and the following attributes:

- quality being part of organizational culture;
- supportive leadership;
- a performance measurement system that supports the quality program;
- employee involvement encouraged;
- quality program linked to strategy; and
- transparency of information regarding quality program.

The following list details the results of the chi square test showing only those variables which displayed the most significant relationships ($p = 0.000$):

- quality is embedded into the organizations culture;
- Key Performance Indicators (KPIs) are identified as part of the strategic planning process;
- operational performance measures link operational activities to the strategic plan;
- quality goals are able to be translated into operational goals;
- the organization environment is such that what gets said gets done;
- employees are encouraged to work smarter not harder;
- managers support staff not by punishing mistakes but by encouraging staff to learn;
- top management is committed to the quality program;
- management ensure that employees are aware of what quality means to the organization;
- all appropriate management and employees are made aware of the performance measures to encourage ongoing improvement;
- performance goals – clear and consistent;
- performance goals – reflect the importance of the quality improvement activities;
- feedback gained from assessing performance against target enables the instigation of rapid corrective action; and
- new ideas are encouraged.

Discussion

The objective of this study was to identify the role of organizational learning mechanisms to support continuous improvement. In particular, a focus was given to the performance measurement system. The findings suggest that for a quality program to be successful the organization must have in place the commitment to learning and adopt supportive control systems flexible enough to meet the changing needs of the business environment. The absence of such attributes may account for the lack of success of some quality programs. As noted by Berling (2000):

> [. . .] the task is not only to start the improvement process, but also to sustain it and to incorporate it into the normal part of everyday work.

In line with Evans and Lindsay (1996), Kaye and Dyason (1995), Kossoff (1993), Lillrank *et al.* (2001), Melan (1993), the findings indicate that quality has been embraced by respondents' organizations as a key competitive weapon to enable survival in the competitive market place. As noted by Butz (1995), Ehrenberg and Stupak (1994), and Bessant and Francis (1999) continuous improvement must be embodied within the strategy if it is to achieve the desired outcomes. Respondents identified that their organizations have focused their strategy on meeting customer needs and improving business processes, both of which suggest a strategy built on continuous improvement. This will be particularly important for a number of respondents' organizations, especially for the 42.3 percent of respondents who consider their organization's product/service is similar or inferior to competitors. This indicates that an opportunity exists for such organizations to improve their operations and close the performance gap thereby improving their position in the market place.

The findings suggest that the quality program has assisted organizations in meeting the strategic objectives. For the majority of respondents (82.7 percent) performance has improved, with 75.8 percent of respondents noting an overall improvement in their organization's competitive position in the market. The majority of respondents (80.8 percent) did not consider that the quality program has had an adverse affect on profitability.

Those organizations that have achieved or exceeded the desired quality outcomes have embedded quality into the organization's culture. This has been assisted by management ensuring that all employees are made aware of what quality means to the organization and by showing their own commitment to the quality initiative. It is interesting to note that respondents who rate their organization's quality initiative as less successful ("fell short of expectations") comment on the lack of employee commitment to the quality program. As noted in the literature, the loss of morale or an increase in cynicism among employees is a major factor in TQM failures (Beck and Yeager, 1996; Dooley and O'Sullivan, 1999).

Therefore, respondents who perceive their organization's quality program as more successful consider that their organization has a culture that encourages both continuous improvement and learning. The learning environment is supported by a performance measurement system that assists in building the shared vision which is necessary for the learning organization (Senge, 1990). Performance measures are clear and consistent and support the strategic objectives of customer satisfaction. The relevance of the performance goals is strengthened by employee involvement in the goal setting process. Performance goals play an important role in assessing operational activities and employee performance. It could be argued that such organizations have become a learning community, whereby, as the individuals learn, the organization learns its way forward.

References

Abraham, M., Crawford, J. and Fisher, T. (1997), "Quality culture and the management of organization change", *International Journal of Quality & Reliability Management*, Vol. 14 No. 6, pp. 616-36.

Ahmed, P.K., Loh, A.Y. and Zairi, M. (1999), "Cultures for continuous improvement and learning", *Total Quality Management*, Vol. 10 No. 4&5, pp. S426-34.

Appelbaum, S.H. and Reichart, W. (1997), "How to measure an organization's learning ability: a learning orientation: part I", *Journal of Workplace Learning*, Vol. 9 No. 7, pp. 225-39.

Banker, R., Potter, G. and Schroeder, R. (1993), "Manufacturing performance reporting for continuous improvement", *Management International Review*, Vol. 33, pp. 69-81.

Beck, J.D.W. and Yeager, N.M. (1996), "How to prevent teams from failing", *Quality Progress*, Vol. 29, pp. 27-31.

Berling, C. (2000), "Continuous improvement as seen from groups and 'improvement agents'", *Total Quality Management*, Vol. 11 Nos 4/5&6, pp. 484-9.

Bessant, J. and Francis, D. (1999), "Developing strategic continuous improvement capability", *International Journal of Operations & Production Management*, Vol. 19 No. 11, pp. 1106-19.

Butz, H.E. (1995), "Strategic planning: the missing link in TQM", *Quality Progress*, Vol. 28 No. 5, pp. 105-8.

Chapman, R. and Hyland, P. (2000), "Strategy and continuous improvement in small-to-medium Australian manufacturers", *Integrated Manufacturing Systems*, Vol. 11 No. 3, pp. 171-9.

Chapman, R.L., Murray, P.C. and Mellor, R. (1997), "Strategic quality management and financial performance indicators", *International Journal of Quality & Reliability Management*, Vol. 14 No. 4, pp. 432-48.

Chenall, R.H. (1997), "Reliance on manufacturing performance measures, total quality management and organizational performance", *Management Accounting Research*, Vol. 8, pp. 187-206.

Claver, E., Tari, J.J. and Molina, J.F. (2002), "Areas of improvement in certified firms advancing towards TQM", *International Journal of Quality & Reliability Management*, Vol. 19 Nos 8/9, pp. 1014-36.

Czuchry, A.J., Hyder, C., Yasin, M. and Mixon, D. (1997), "A systematic approach to improving quality: a framework and a field study", *International Journal of Quality & Reliability Management*, Vol. 14 No. 9, pp. 876-98.

Dahlgaard, S. (1999), "The evolution patterns of quality management: some reflections on the quality movement", *Total Quality Management*, Vol. 10 No. 4&5, pp. 473-80.

Daniel, S. and Reitsperger, W.D. (1992), "Management control systems for quality: an empirical comparison of the US and Japanese electronics industries", *Journal of Management Accounting Research*, Vol. 4, pp. 64-78.

Dean, J.W. and Bowen, D.E. (1994), "Management theory and total quality: improving research and practice through theory development", *The Academy of Management Review*, Vol. 19 No. 3, pp. 392-418.

Deming, W.E. (1993), *The New Economics for Industry, Government, Education*, Centre for Advanced Engineering Study, Massachusetts Institute of Technology, Cambridge, MA.

DiBella, A. and Nevis, E. (1998), *How Organizations Learn*, Jossey-Bass Publishers, San Fransisco, CA.

Dodgson, M. (1993), "Organizational learning: a review of some literatures", *Organization Studies*, Vol. 14 No. 3, pp. 375-94.

Dooley, L. and O'Sullivan, D. (1999), "Decision support system for the management of systems change", *Technovation*, Vol. 19 No. 8, pp. 483-93.

Dossett, D., Latham, G. and Mitchell, T. (1979), "Effects of assigned versus participatively set goals, knowledge of results, and individual differences on employee behaviour when goal difficulty is held constant", *Journal of Applied Psychology*, Vol. 64 No. 3, pp. 291-8.

Egan, G. (1993), *Adding Value*, Jossey-Bass Publishers, San Francisco, CA.

Ehrenberg, R.H. and Stupak, R.J. (1994), "Total quality management: its relationship to administrative theory and organizational behavior in the public sector", *Public Administration Quarterly*, Vol. 18 No. 1, pp. 75-98.

Evans, J.R. and Lindsay, W.M. (1996), *The Management and Control of Quality*, 3rd ed., West Publishing Company, St Paul, MN.

Fine, C. (1986), "Quality improvement and learning in productive systems", *Management Science*, Vol. 32 No. 10, pp. 1301-15.

Fiol, C.M. and Lyles, M. (1985), "Organizational learning", *Academy of Management Review*, Vol. 10, pp. 803-13.

Groth, J.C. (1995), "Total quality management: perspective for leaders", *The TQM Magazine*, Vol. 7 No. 3, pp. 54-9.

Hendricks, K.B. and Singhal, V.R. (1997), "Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards", *Management Science*, Vol. 43 No. 9, pp. 1258-74.

Hoque, Z. and Alam, M. (1999), "TQM adoption, institutionalism and changes in management accounting systems: a case study", *Accounting and Business Research*, Vol. 29 No. 3, pp. 199-210.

Huber, G.P. (1991), "Organizational learning: the contributing processes and the literatures", *Organization Science*, Vol. 2, pp. 88-115.

Jha, S., Noori, H. and Michela, J.L. (1996), "The dynamics of continuous improvement: aligning organizational attributes and activities for quality and productivity", *International Journal of Quality Science*, Vol. 1 No. 1, pp. 19-47.

Kaye, M. and Dyason, M. (1995), "The fifth era", *The TQM Magazine*, Vol. 7 No. 1, pp. 33-7.

Kloot, L. (1997), "Organizational learning and management control systems: responding to environmental change", *Management Accounting Research*, Vol. 8, pp. 47-73.

Kossoff, L.L. (1993), "Total quality or total chaos?", *HR Magazine*, Vol. 38 No. 4, pp. 131-4.

Krishnan, R., Shani, A.B., Grant, R.M. and Baer, R. (1993), "In search of quality improvement: problems of design and implementation", *Academy of Management Executive*, Vol. 7 No. 4, pp. 7-20.

Lau, R.S.M. and Anderson, C.A. (1998), "A three-dimensional perspective of total quality management", *International Journal of Quality & Reliability Management*, Vol. 15 No. 1, pp. 85-98.

Lee, T.H. (1995), "Learning: what does it mean?", *Center for Quality of Management Journal*, Vol. 4 No. 4, pp. 4-14.

Lee, T.H. and Walden, D. (1998), "Designing integrated management system", *Center for Quality of Management Journal*, Vol. 7 No. 1, pp. 3-18.

Liedtka, J.M. and Rosenblum, J.W. (1996), "Shaping conversations: making strategy, managing change", *California Management Review*, Vol. 39 No. 1, pp. 141-57.

Lillrank, P., Shani, A.B. and Lindberg, P. (2001), "Continuous improvement: exploring alternative organizational designs", *Total Quality Management*, Vol. 12 No. 1, pp. 41-55.

Lipshitz, P.M. and Oz, S. (1996), "Building learning organizations: the design and implementation of organizational learning mechanisms", *Journal of Applied Behavioral Science*, Vol. 32 No. 3, pp. 292-305.

Lipshitz, R. and Popper, M. (2000), "Organizational learning in a hospital", *Journal of Applied Behavioral Science*, Vol. 36 No. 3, pp. 345-61.

Llopis, J. and Tari, J.J. (2001), "The importance of internal aspects of quality improvement", *International Journal of Quality & Reliability Management*, Vol. 20 No. 3, pp. 304-24.

Locke, E.A., Saari, L.M., Shaw, K.N. and Latham, G.P. (1981), "Goal setting and task performance: 1969-1980", *Psychological Bulletin*, Vol. 90 No. 1, pp. 125-52.

Lorente, A.R., Dewhurst, F. and Dale, B.G. (1999), "TQM and business innovation", *European Journal of Innovation Management*, Vol. 2 No. 1, pp. 12-19.

Lowe, E.A. (1971), "On the idea of a management control system", *Journal of Management Studies*, February, pp. 1-12.

Melan, E. (1993), "Quality improvement in higher education: TQM in administrative functions", *CUPA Journal*, Vol. 44, pp. 7-18.

Oakland, J.S. (1993), *Total Quality Management*, 2nd ed., Heinemann, London.

Reeves, C. and Bednar, D. (1994), "Defining quality: alternatives and implications", *The Academy of Management Review*, July, pp. 419-37.

Sambrook, S. and Stewart, J. (2000), "Factors influencing learning in European learning oriented organizations: issues for management", *Journal of European Industrial Training*, Vol. 24 Nos 2/3/4, pp. 209-19.

Senge, P.M. (1990), *The Fifth Discipline*, Doubleday, New York, NY.

Simons, R. (1990), "The role of management control systems in creating competitive advantage: new perspectives", *Accounting, Organizations and Society*, Vol. 15 Nos 1/2, pp. 127-43.

Simons, R. (1991), "Strategic orientation and top management attention to control systems", *Strategic Management Journal*, Vol. 12, pp. 49-62.

Simons, R. (1992), "The strategy of control", *CA Magazine*, March, pp. 44-6.

Sinclair, D. and Zairi, M. (1995), "Performance measurement as an obstacle to TQM", *The TQM Magazine*, Vol. 7 No. 2, pp. 42-5.

Sitkin, S.B., Sutcliffe, K.M. and Schroeder, R.G. (1994), "Distinguishing control from learning in total quality management: a contingency perspective", *The Academy of Management Review*, Vol. 19 No. 3, pp. 537-57.

Spong, G. (1994), "A quality approach", *Australian Small Business and Investing*, July, p. 80.

Srinidhi, B. (1998), "Strategic quality management", *International Journal of Quality Science*, Vol. 3 No. 1, pp. 38-70.

Terziovski, M., Howell, A., Sohal, A. and Morrison, M. (2000), "Establishing mutual dependence between TQM and the learning organization: a multiple case study analysis", *The Learning Organization*, Vol. 7 No. 1, pp. 23-31.

Tranfield, D., Duberley, J., Smith, S., Musson, G. and Stokes, P. (2000), "Organizational learning – it's just routine", *Management Decision*, Vol. 38 No. 4, pp. 253-60.

Waldman, D.A. (1994), "Designing performance management systems for total quality implementation", *Journal of Organizational Change Management*, Vol. 7 No. 2, pp. 31-44.

Wick, C.W. and León, L.S. (1995), "From ideas to action: creating a learning organization", *Human Resource Management*, Vol. 34 No. 2, pp. 299-311.

Wright, P.M. (1994), "Goal setting and monetary incentives: motivational tools that can work too well", *Compensation and Benefits Review*, May/June, pp. 41-9.

Yong, J. and Wilkinson, A. (2001), "Rethinking total quality management", *Total Quality Management*, Vol. 12 No. 2, pp. 247-58.

Further reading

Bessant, J. and Caffyn, S. (1997), "High-involvement innovation through continuous improvement", *International Journal of Technology Management*, Vol. 14 No. 1, pp. 7-22.

Simons, R. (1995), *Levers of Control: How Managers Use Innovative Control Systems to Drive Strategic Renewal*, Harvard Business School Press, Boston, MA.

About the author

Judy Oliver is a senior lecturer in accounting at Swinburne University. She has previously held appointments at Victoria University and University of Tasmania. Her teaching has been in the area of management accounting and introductory financial accounting. Her research interests lie in management accounting, total quality management and corporate governance. Judy Oliver can be contacted at: juditholiver@swinburn.edu.au

IJQRM
26,6

Received 18 May 2008
Revised 15 December 2008
Accepted 19 December 2008

The effect of gauge measurement capability on MC_p and its statistical properties

Davood Shishebori and Ali Zeinal Hamadani
Department of Industrial Engineering, Isfahan University of Technology, Isfahan, Iran

Abstract

Purpose – The aim of this paper is to consider the effect of gauge measurement capability on the multivariate process capability index (MC_p).

Design/methodology/approach – With respect to measurement capability, the paper investigates the statistical properties of the estimated MC_p and considers the effect of gauge measurement capability on the lower confidence bound, hypothesis testing, critical value and power of testing for MC_p at the mentioned state.

Findings – The results show that gauge measurement capabilities will notably change the results of estimating and testing the process capability index.

Originality/value – The research would help quality experts to determine whether their processes meet the required capability, and to make more reliable decisions.

Keywords Process analysis, Confidence limits, Multivariate analysis, Manufacturing industries, Performance measures

Paper type Research paper

Introduction

In the past few years, process capability analysis have been introduced and used to characterize process performance. This analysis essentially aims at checking the adequacy of a process to meet the specification limits. Also, it is a measure of the inherent precision of a manufacturing process. One of the methods for process capability analysis is indices. The measurement and evaluation of the production process performance using process capability indices (PCI) are becoming more frequent. Process capability indices establish the relationships between the actual process performance and the manufacturing specifications and provide an effective measure of process performance. Chen and Cherng (2001) mentioned that many process capability indices have been proposed to provide numerical measures on process performance. Engineers can realize whether the product meets its specifications using process capability indices. These capability indices have been the focus of recent research in quality assurance and process capability analysis and have been widely used in the manufacturing industry for measuring process reproduction capability according to its manufacturing specifications. These indices, quantifying process potential and performance, are important for any successful quality improvement activities and quality program implementation. In current practice, suppliers are often required to provide their process capability of the product to the customers in the

International Journal of Quality & Reliability Management
Vol. 26 No. 6, 2009
pp. 564-582

0265-671X
DOI 10.1108/02656710910966138

supply chain partnership. Process capability indices can also be used as the benchmarking for quality improvement activities. Capability indices, C_p, C_{pk}, C_{pm}, and C_{pmk}, have been proposed to evaluate process performance but restricted to cases with single engineering specification. A large number of papers have dealt with the statistical properties and the estimation of these univariate indices. Kotz and Johnson (2002) provided a compact survey and commented on some 170 publications on process capability indices during the years 1992 to 2000. Also, Pearn and Kotz (2006) provided a comprehensive survey on process capability indices during the beginning of introducing these indices up to late 2005.

One interesting fact about the characteristic measuring in a process is that the inevitable variations in process measurements come from two sources: the manufacturing process and the gauge. Gauge capability reflects the gauge's precision, or lack of variation, but is not the same as calibration which assures the gauge's accuracy. As it has been emphasized in the numerous occasions, process capability measures the ability of a manufacturing process to meet reassigned specifications. Nowadays, many customers use process capability to judge supplier's ability to deliver quality products. Suppliers ought to be aware of how gauges affect various process capability estimates.

However, no measurement is free from error or uncertainty even though it may be conducted with the aid of highly sophisticated measuring instruments. Montgomery and Runger (1993) pointed out that the quality of data on the process characteristics relies very much on the gauge. Pearn *et al.* (2007) mentioned that any variation in the measurement process has a direct impact on the ability to make sound judgment about the manufacturing process. An inaccurate measurement system can thwart all the benefits of improvement endeavours resulting in poor quality. Analyzing process capability without considering gauge capability may often lead to unreliable decisions. It could result in a serious loss to producers if gauge capability is not being considered in process capability estimation and testing. On the other hand, improving the gauge measurements and employing properly trained operators can reduce the measurement errors. Since measurement errors unfortunately cannot be avoided, using appropriate confidence coefficients and power becomes an essential task. However, the reality is that no measurement is free from error or uncertainty even if it is carried out with the aid of highly sophisticated and precise measuring instruments.

An extensive study has been done for the case of univariate process capability with considering the measurement error such as analyzing the effects of measurement errors on process capability indices, Mittag (1997) and Levinson (1995) provided Some very definitive techniques for quantifying the percentage error in process capability indices estimation in the presence of measurement errors. Pearn and Kotz (2006) provided a compact survey on pervious researches process capability indices with gauge measurement errors during the beginning of introducing these indices up to late 2005.

Another point that is considered in process capability indices and found out from the history of this subject; is the bulk of the studies associated with analyzing the quality and efficiency of a process are so far limited to discussing one single quality specification but in real applications, manufactured products often have multiple quality characteristics and multiple characteristics processes are by now so common that our studies to capability indices cannot be restricted to the univariate domain. That is, the process capability analysis involves more than one engineering specification. For this reason, multivariate methods for assessing process capability are proposed.

Chan *et al.* (1991), Taam *et al.* (1993), Pearn *et al.* (1992), Chen (1994), Karl *et al.* (1994), Shahriari *et al.* (1995), Boyles (1996), Wang and Du (2000), Wang *et al.* (2000), and others have developed and presented multivariate capability indices for assessing capability. Wang and Chen (1998) and Wang and Du (2000) proposed multivariate extensions for C_p, C_{pk}, C_{pm}, and C_{pmk} based on the principal component analysis, which transforms numbers of original related measurement variables into a set of uncorrected linear functions. A comparison of three novel multivariate methodologies for assessing capability is illustrated in Wang *et al.* (2000). Although some multivariate capability indices have been studied and an extensive study has been done for the case of univariate process capability with considering the measurement error; but there is a real need for considering this effect on the multivariate quality characteristics and there is not any study in this ground.

In this paper we introduce the common capability index, C_p in multivariate state by considering gauge measurement capability and consider the effect of measurement errors on the C_p and its statistical properties. Since gauge measurement errors has an negative effect on the estimating and testing of MC_p; so we present adjusted confidence interval bounds and critical values for capability testing purpose of MC_p with unavoidable measurement errors.

Multivariate process capability in presence of gauge measurement errors

In this paper, we will focus on the multivariate process index MC_p (Taam *et al.*, 1993). The multivariate capability index MC_p is defined as:

$$\begin{aligned} MC_p &= \frac{\text{vol(modified tolerance region)}}{\text{vol}[(X-\mu)'\Sigma^{-1}(X-\mu) \leq k(q)]} \\ &= \frac{\text{vol(modified tolerance region)}}{(\pi\chi^2_{v,0.9973})^{v/2}|\Sigma|^{1/2}\{\Gamma[(v/2)+1]\}^{-1}}, \end{aligned} \tag{1}$$

where $k(q)$ is the 99.73th percentile of the χ^2 distribution with v degrees of freedom or the dimension of variables, μ is the mean vector and Σ represents the variance-covariance matrix of X, $|\Sigma|$ is the determinant of Σ and $\Gamma(.)$ is the gamma function.

Gauge repeatability and reproducibility (GR&R) studies focus on quantifying the measurement errors. Common approaches to GR&R studies, such as the Range method (Montgomery and Runger, 1993) assume that the distribution of the measurement errors is normally distributed with a mean error of zero. Suppose that in multivariate state, the measurement errors are described by a random variable $M \sim \text{Normal}(\mu_{\text{Me}}, \Sigma_{\text{Me}})$, where $\mu_{\text{Me}} = 0$ is the mean vector and Σ_{Me} is the variance covariance matrix of the measurement error. So based on definition of Montgomery and Runger (1993), the gauge capability at the multivariate state defined by:

$$\lambda^M = \frac{\mathrm{vol}[(X-\mu_{\mathrm{Me}})'\Sigma_{\mathrm{Me}}^{-1}(X-\mu_{\mathrm{Me}}) \le k(q)]}{\mathrm{vol(modified\ tolerance\ region)}}$$

$$= \frac{(\pi\chi^2_{v,0.9973})^{v/2}|\Sigma_{\mathrm{Me}}|^{1/2}\{\Gamma[(v/2)+1]\}^{-1}}{\mathrm{vol(modified\ tolerance\ region)}}, \quad (2)$$

where $|\Sigma_{\mathrm{Me}}|$ is the determinant of Σ_{Me}, μ_{Me} is the mean vector of errors and χ^M is the gauge capability index. For the measurement system to be deemed acceptable, the variability in the measurements due to the measurement system must be less than a predetermined percentage of the engineering tolerance. So based on the recommendation of the Automotive Industry Action Groups, some guidelines for gauge acceptance offered (Montgomery, 1996).

Considering the process capability in the measurement error system, we assume that the observations X have a multivariate normal distribution $N_v(\mu,\Sigma)$ and show the relevant quality characteristic of a manufacturing process. Because of measurement errors, the observed variable $Y \sim \mathrm{N_v}$ ($\mu_Y = \mu$, $\Sigma_Y = \Sigma + \Sigma_{\mathrm{Me}}$) is measured by the assumption that X and Me are stochastically independent, instead of measuring the true variable X. The empirical process capability index (MC_p^Y) is obtained after substituting Σ_Y for Σ, so the multivariate capability index MC_p^Y is defined as:

$$MC_p^Y = \frac{\mathrm{vol(modified\ tolerance\ region)}}{\mathrm{vol}[(X-\mu)'\Sigma_Y^{-1}(X-\mu) \le k(q)]}$$

$$= \frac{\mathrm{vol(modified\ tolerance\ region)}}{(\pi\chi^2_{v,0.9973})^{v/2}|\Sigma_Y|^{1/2}\{\Gamma[(v/2)+1]\}^{-1}}. \quad (3)$$

It is easy to show that the relationship between the true process MC_p and the empirical process capability MC_p^Y is given as (Shishebori and Hamadani, 2008):

$$MC_p^Y = \frac{MC_p}{\sqrt{(\lambda^M MC_p)^2 + \frac{|\Sigma_Y|-|\Sigma_{Me}|}{|\Sigma_Y - \Sigma_{Me}|}}}. \quad (4)$$

Since the variation of data we observe is larger than that of the variation of the original data, the denominator of the index MC_p becomes larger and we will underestimate the true capability of the process.

Estimation of MC_p in presence of gauge measurement errors

An estimator of MC_p can be expressed as:

$$\hat{M}C_p = \frac{\mathrm{vol(modified\ tolerance\ region)}}{\mathrm{vol(estimated\ 99.73\ per\ cent\ process\ region)}}$$

$$= \frac{\mathrm{vol(modified\ tolerance\ region)}}{(\pi\chi^2_{v,0.9973})^{v/2}|S|^{1/2}\{\Gamma[(v/2)+1)\}^{-1}},$$

where S is the sample variance-covariance matrix from process and $|S|$ is the determinant of S.

$\hat{MC}_p$ is a biased estimator of MC_p and multiplied by b_v, given as:

$$b_v = \left(\frac{2}{n-1}\right)^{v/2} \frac{\Gamma[1/2(n-1)]}{\Gamma[1/2(n-v)-1/2]}.$$

We get an unbiased estimation of MC_p as $\tilde{MC}_p = b_v \hat{MC}_p$. Pearn *et al.* (2007) showed that $\tilde{MC}_p$ is the UMVUE (uniformly minimum variance unbiased estimator) of MC_p.

With respect to gauge measurement capability and using the estimators, S_Y, S_{Me} and $\hat{MC}_p$ for the parameters Σ_Y, Σ_{Me} and MC_p, the biased estimator of MC_p is given as:

$$\begin{aligned}\hat{MC}_p^Y &= \frac{\text{vol(modified tolerance region)}}{(\pi\chi^2_{v,0.9973})^{v/2}|S_Y|^{1/2}\{\Gamma[(v/2)+1]\}^{-1}} \\ &= \frac{\text{vol(modified tolerance region)}}{(\pi\chi^2_{v,0.9973})^{v/2}|S+S_{\mathrm{Me}}|^{1/2}\{\Gamma[(v/2)+1)]\}^{-1}},\end{aligned} \tag{5}$$

where S_Y is the sample variance-covariance matrix and $|S_Y|$ is the determinant of S_Y.

So the relationship between the estimators of the true process MC_p and the empirical process capability MC_p^Y is given as:

$$\hat{MC}_p^Y = \frac{\hat{MC}_p}{\sqrt{(\hat{\lambda}^M \hat{MC}_p)^2 + \frac{|S_Y|-|S_{\mathrm{Me}}|}{|S_Y - S_{\mathrm{Me}}|}}}, \text{ or } \tilde{MC}_p^Y = \frac{\tilde{MC}_p}{\sqrt{(\tilde{\lambda}^M \tilde{MC}_p)^2 + \frac{|S_Y|-|S_{\mathrm{Me}}|}{|S_Y - S_{\mathrm{Me}}|}}}. \tag{6}$$

Illustration example

Chen (1994) discussed a bivariate normal example and employed Sultan (1986) bivariate process data ($n = 25$) of the Brinell hardness (H) and the tensile strength (S) of a particular process. The specification limits for H and S were set at $(112.7, 241.3)$ and $(32.7, 73.3)$, respectively. The centre of the specifications was $\mu_0^T = [177,\ 53]$. The sample mean vector and sample covariance matrix were:

$$\bar{X}^T = [177.2,\ 52.32],$$

$$S_Y = \begin{bmatrix} 348.8347 & 85.3308 \\ 85.3308 & 44.6594 \end{bmatrix}.$$

Using $\chi^2_{2,0.9973} = 11.829$ and $|S_Y| = 8,297.4$, then we obtain the practical estimated value of process capability index to be:

$$\hat{MC}_p^Y = \frac{\pi \times [(241.3-112.7)/2] \times [(73.3-32.7)/2]}{(\pi \times \chi^2_{2,\,0.9973})|S_Y|^{1/2}} = 1.2114,$$

and

$$\tilde{M}C_p^Y = (11/12) \times (1.2114) = 1.1104.$$

This value is calculated by ignoring the gauge measurement capability. Now by considering the gauge measurement error for the data, and with respect to the independence of measuring instruments for two variables; we assume that the variance-covariance matrix of gauge measurement is:

$$S_{\mathrm{Me}} = \begin{bmatrix} 11.0347 & 0 \\ 0 & 11.0347 \end{bmatrix}.$$

Thus, we get $|S_{\mathrm{Me}}| = 121.7637$ as an estimation of $|\Sigma_{\mathrm{Me}}|$. Using (2), one can get the gauge measurement capability as 0.1 ($\lambda^M = 0.1$); therefore, from (6) we obtain $\hat{M}C_p = 1.7282$ and $\tilde{M}C_p = 1.5842$.

Comparing $\hat{M}C_p(\tilde{M}C_p)$ with $(\hat{M}C_p^Y(\tilde{M}C_p^Y)$, it is obvious that the effect of λ^M on MC_p^Y will increase; in other words, with increasing λ^M, MC_p^Y will decrease. Figure 1 shows the estimates of MC_p^Y for different values of λ^M using $\hat{M}C_p = 2.9208$.

As one can see in Figure 1, by increasing MC_p, the effect of λ^M on MC_p^Y will be increased; in other words, with increasing λ^M; MC_p^Y will be decreased.

Expected value, variance and MSE of $\hat{M}C_p^Y$

According to Pearn *et al.* (2007), the probability density function of $\hat{M}C_p^Y$ is expressed as:

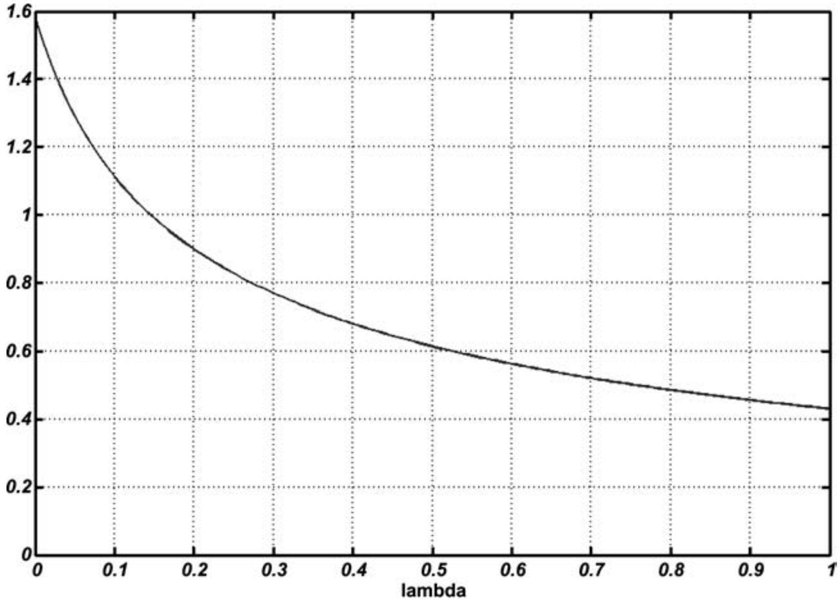

Figure 1. The variation of MC_p^Y with $\lambda^M = [0, 1]$ for $MC_p = 1.5842$

$$f(x) = f_Y[g^{-1}(x)] \times \left| \frac{d}{dx} g^{-1}(x) \right|$$

$$= f_Y \left[\frac{(MC_p^Y)^2(n-1)^v}{x^2} \right] \times \frac{2(MC_p^Y)^2(n-1)^v}{x^3} \text{ for } x > 0. \quad (7)$$

Since the hth moment of a χ^2 distribution with v degrees of freedom is $2^h\Gamma(v/2+h)/\Gamma(v/2)$, using the fact that the moment of product of independent variables is the product of the moments of the variables, the hth moment of $|S_Y|/\Sigma_Y$ can be obtained as:

$$E(|S_Y|/\Sigma_Y)^h = \frac{2^{vh}\prod_{i=1}^{v}\Gamma[1/2(n-i)+h]}{(n-1)^{vh}\prod_{i=1}^{v}\Gamma[1/2(n-i)]}. \quad (8)$$

Now we can derive the rth moment of $\hat{M}C_p^Y$ according to equation (7) and $\hat{M}C_p^Y = MC_p^Y \times (|S_Y|/\Sigma_Y)^{-1/2}$, the result is given by:

$$E[(\hat{M}C_p^Y)^r] = E[(MC_p^Y)^r \times (|SY|/\Sigma_Y)^{-r/2}] = (MC_p^Y)^r \times E(|S_Y|/\Sigma_Y)^{-r/2}. \quad (9)$$

Now, by substituting $r = 1$ and $h = -1/2$ into equations (8) and (9), respectively, we have:

$$E(\hat{M}C_p^Y) = MC_p^Y \times \frac{2^{-v/2}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1/2]}{(n-1)^{-v/2}\prod_{i=1}^{v}\Gamma[1/2(n-i)]}$$

$$= MC_p^Y \times \left(\frac{n-1}{2}\right)^{v/2} \frac{\Gamma[1/2(n-v)-1/2]}{\Gamma[1/2(n-1)]} = \frac{1}{b_v} \times MC_p^Y, \quad (10)$$

where b_v is a correction factor, so that $\tilde{M}C_p^Y = b_v \times \hat{M}C_p^Y$ is an unbiased estimator of MC_p^Y. Substitute $r = 2$ and $h = -1$ into equations (8) and (9), we get:

$$E[(\hat{M}C_p^Y)^2] = (MC_p^Y)^2 \times \frac{2^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1]}{(n-1)^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)]}. \quad (11)$$

From equations (10) and (11), we have the variance of $\hat{M}C_p^Y$ as:

$$\mathrm{Var}(\hat{M}C_p^Y) = (MC_p^Y)^2 \times \frac{2^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1]}{(n-1)^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)]} - \left(\frac{MC_p^Y}{b_v}\right)^2$$

$$= \left[\frac{2^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1]}{(n-1)^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)]} - \left(\frac{1}{b_v}\right)^2\right]\frac{(MC_p)^2}{(\lambda^M MC_p)^2 + \frac{|\Sigma_Y|-|\Sigma_{\mathrm{Me}}|}{|\Sigma_Y-\Sigma_{\mathrm{Me}}|}}. \quad (12)$$

Since n is a finite positive integer:

$$\left[2^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1]/(n-1)^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)] - (1/b_v)^2\right]$$

is positive; so we have $\mathrm{Var}(\hat{M}C_p^Y) < \mathrm{Var}(\hat{M}C_p)$. For $\lambda^M > 0$ it is clear that $\tilde{M}C_p^Y$ is a biased estimator of MC_p, and the bias is given as:

$$\mathrm{bias}(\tilde{M}C_p^Y) = E(\tilde{M}C_p^Y) - MC_p = \left(\frac{1}{\sqrt{(\lambda^M MC_p)^2 + \frac{|\Sigma_Y|-|\Sigma_{\mathrm{Me}}|}{|\Sigma_Y-\Sigma_{\mathrm{Me}}|}}} - 1\right)MC_p, \quad (13)$$

which is a decreasing function of λ^M.

Taking into account both the bias and the variance of the estimators $\tilde{M}C_p$ and $\tilde{M}C_p^Y$. Using the fact that MSE $=$ (bias)^2 $+$ variance, the MSEs of $\tilde{M}C_p$ and $\tilde{M}C_p^Y$, denoted by $\mathrm{MSE}(\tilde{M}C_p)$ and $\mathrm{MSE}(\tilde{M}C_p^Y)$ are given as:

$$\mathrm{MSE}(\tilde{M}C_p) = MC_p^2 \times \left[b_v^2 \frac{2^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1]}{(n-1)^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)]} - 1\right], \quad (14)$$

$$\mathrm{MSE}(\tilde{M}C_p^Y) = MC_p^2 \times \left[1 + b_v^2 \frac{2^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)-1]}{(n-1)^{-v}\prod_{i=1}^{v}\Gamma[1/2(n-i)]} \times \frac{1}{(\lambda^M MC_p)^2 + \frac{|\Sigma_Y|-|\Sigma_{\mathrm{Me}}|}{|\Sigma_Y-\Sigma_{\mathrm{Me}}|}} + \left(\frac{2}{\sqrt{(\lambda^M MC_p)^2 + \frac{|\Sigma_Y|-|\Sigma_{\mathrm{Me}}|}{|\Sigma_Y-\Sigma_{\mathrm{Me}}|}}}\right)\right]. \quad (15)$$

For comparing MSE($\tilde{M}C_p^Y$) with (MSE($\tilde{M}C_p$)), we consider the function:

$$\Gamma = G(MC_p, n, \lambda^M) = \frac{\mathrm{MSE}(\tilde{M}C_p^Y)}{\mathrm{MSE}(\tilde{M}C_p)} = \frac{[\mathrm{bias}(\tilde{M}C_p^Y)]^2 + \mathrm{variance}(\tilde{M}C_p^Y)}{[\mathrm{bias}(\tilde{M}C_p)]^2 + \mathrm{variance}(\tilde{M}C_p)}. \tag{16}$$

Figure 2 shows the three dimensional surface of Γ for $MC_p = 1.5842$ and different values of sample size (n) and λ^M.

Confidence interval for MC_p

Since $\tilde{M}C_p$ is an statistical estimator like other statistics, it is subject to the sampling variation, therefore one needs to compute an interval to provide a range that includes the true MC_p with high probability. Based on the definition, a $100(1-\alpha)$ percent confidence interval for MC_p can be established as (Pearn *et al.*, 2007) $100(1-\alpha)$ percent confidence interval bound can be written as (17):

$$\left[(\hat{M}C_p)^2 \frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}{(n-1)^v}, (\hat{M}C_p)^2 \frac{F_Y^{-1}\left(\frac{\alpha}{2}\right)}{(n-1)^v}\right]$$
$$\text{or} \left[\left(\frac{\tilde{M}C_p}{b_v}\right)^2 \frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}{(n-1)^v}, \left(\frac{\tilde{M}C_p}{b_v}\right)^2 \frac{F_Y^{-1}\left(\frac{\alpha}{2}\right)}{(n-1)^v}\right]. \tag{17}$$

Furthermore, a $100(1-\alpha)$ percent lower confidence bound for MC_p can be obtained as:

$$\left[(\hat{M}C_p)^2 \frac{F_Y^{-1}(\alpha)}{(n-1)^v}\right] \text{or} \left[\left(\frac{\tilde{M}C_p}{b_v}\right)^2 \frac{F_Y^{-1}(\alpha)}{(n-1)^v}\right]. \tag{18}$$

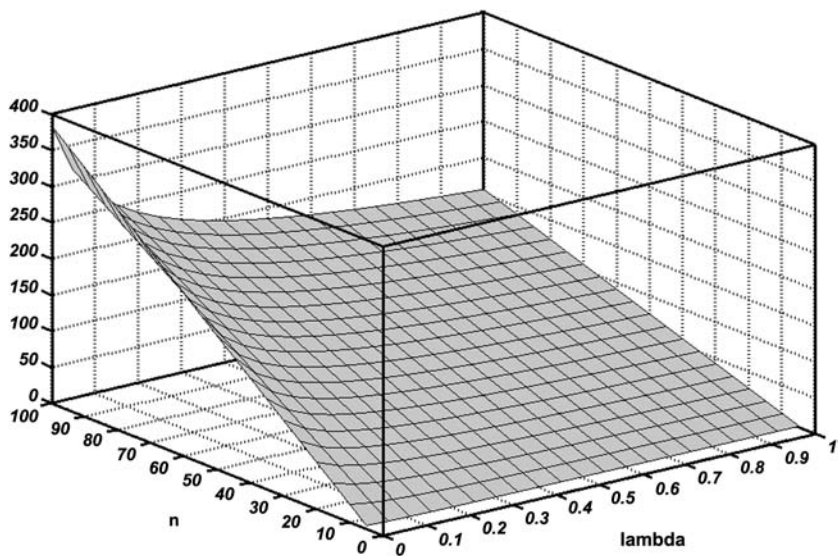

Figure 2.
Surface plot of γ with $n = 5(5)100$ and λ^M in $[0, 1]$ for $MC_p = 1.5842$

However, as a result of the measurement errors, we take $\tilde{M}C_p^Y$ as an estimator of MC_p. Thus the confidence bounds are:

$$\left[(\hat{M}C_p^Y)^2\frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}{(n-1)^v},(\hat{M}C_p^Y)^2\frac{F_Y^{-1}\left(\frac{\alpha}{2}\right)}{(n-1)^v}\right]\text{ or }\left[\left(\frac{\hat{M}C_p^Y}{b_v}\right)^2\frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}{(n-1)^v},\left(\frac{\tilde{M}C_p^Y}{b_v}\right)^2\frac{F_Y^{-1}\left(\frac{\alpha}{2}\right)}{(n-1)^v}\right], \tag{19}$$

$$\left[(\hat{M}C_p^Y)^2\frac{F_Y^{-1}(1-\alpha)}{(n-1)^v},\infty\right]\text{ or }\left[\left(\frac{\tilde{M}C_p^Y}{b_v}\right)^2\frac{F_Y^{-1}(1-\alpha)}{(n-1)^v},\infty\right]. \tag{20}$$

In example discussed by Chen (1994), a 95 percent confidence interval and lower bound for MC_p is given as $[0.5416, 2.2865]$ and $[0.6295, \infty)$.

It is interesting to find out the confidence coefficient, θ (the probability that the confidence interval contains the actual MC_p value) for the confidence bound given in Equation (19). One can calculate this coefficient using the following definition:

$$\begin{aligned}\theta &= P\left((\hat{M}C_p^Y)^2\frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}{(n-1)^v}\le MC_p\le(\hat{M}C_p^Y)^2\frac{F_Y^{-1}\left(\frac{\alpha}{2}\right)}{(n-1)^v}\right)\\ &= P\left(\frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}{(\hat{\lambda}^M\hat{M}C_p)^2+\frac{|S_Y|-|S_{\text{Me}}|}{|S_Y-S_{\text{Me}}|}}\le y\le\frac{F_Y^{-1}\left(\frac{\alpha}{2}\right)}{(\hat{\lambda}^M\hat{M}C_p)^2-\frac{|S_Y|-|S_{\text{Me}}|}{|S_Y-S_{\text{Me}}|}}\right).\end{aligned} \tag{21}$$

By substituting y in the above equation we get:

$$\left(\frac{MC_p^Y}{\hat{M}C_p^Y}\right)^2\frac{y}{(n-1)^v}\Rightarrow\tilde{y}\frac{\hat{M}C_p^2(n-1)^v}{(\hat{M}C_p^Y)^2\left[(\hat{\lambda}^M\hat{M}C_p)^2+\frac{|S_Y|-|S_{\text{Me}}|}{|S_Y-S_{\text{Me}}|}\right]}.$$

If we are interested in evaluating θ for the example given by Chen (1994), then $\theta = 0.5560$. In other words, the probability that the calculated confidence interval contains the real value of MC_p is equal to 0.5560, which is small compared to 0.95. Accordingly, producers will be damaged if they ignore the effect of gauge measurement error on the calculation of confidence interval which will result in rejecting many of their conformed products and making a lot of losses for their process.

In order to improve the confidence interval for the given confidence coefficient ($\theta = 1-\alpha$), one can recalculate the confidence bounds such that it contains the actual value of MC_p with the probability of θ. Hence if we consider the proposed confidence interval to be L^* and U^* then with respect to the gauge measurement capability, the adjusted $100(1-\alpha)$ percent confidence interval bound can be written as (22) and (23):

$$L^*=\sqrt{\frac{F_Y^{-1}\left(1-\frac{\alpha}{2}\right)\times(\hat{M}C_p^Y)^2\times\left(\frac{|S_Y|-|S_{\text{Me}}|}{|S_Y-S_{\text{Me}}|}\right)}{(n-1)^v-(\hat{\lambda}^M\hat{M}C_p^Y)^2F_Y^{-1}\left(1-\frac{\alpha}{2}\right)}}, \tag{22}$$

$$U^* = \sqrt{\frac{F_Y^{-1}\left(\frac{\alpha}{2}\right) \times (\hat{M}C_p^Y)^2 \times \left(\frac{|S_Y|-|S_{\mathrm{Me}}|}{|S_Y - S_{\mathrm{Me}}|}\right)}{(n-1)^v - (\hat{\lambda}^M \hat{M}C_p^Y)^2 F_Y^{-1}\left(\frac{\alpha}{2}\right)}}. \tag{23}$$

For the Chen example, the new confidence interval is given as $L^* = 1.03073$, $U^* = 2.4673$. Therefore, the 0.95 confidence interval for the actual value of MC_p is $[1.0307, 2.4673]$.

In Figure 3, the changing pattern of the 95 percent confidence coefficient as a function of the gauge measurement capability is shown for different sample size (n).

According to the Figure 3, for a fixed value of sample size, the confidence coefficient (θ) will decrease considerably by increasing the measurement capability index. In other words the probability that calculated 95 percent confidence interval, contains the actual value of the MC_p will be reduced. The adjusted confidence interval given by (22), (23) will now contains the actual value of the MC_p with the exact probability of given $\theta = 1 - \alpha$.

Figure 4 shows the curve of the unadjusted lower confidence bound as a function of the measurement capability index for $\alpha = 0.05$ and in Figure 5 one can see the difference between the adjusted and unadjusted lower confidence bound for the same parameters. In both figures the upper straight line show the lower confidence bound for the case of no measurement error.

Hypothesis testing for capability index under gauge measurement errors

In hypothesis testing, we determine whether or not a hypothesized value of a parameter is true or not, based on the sample taken and the parameter estimate derived

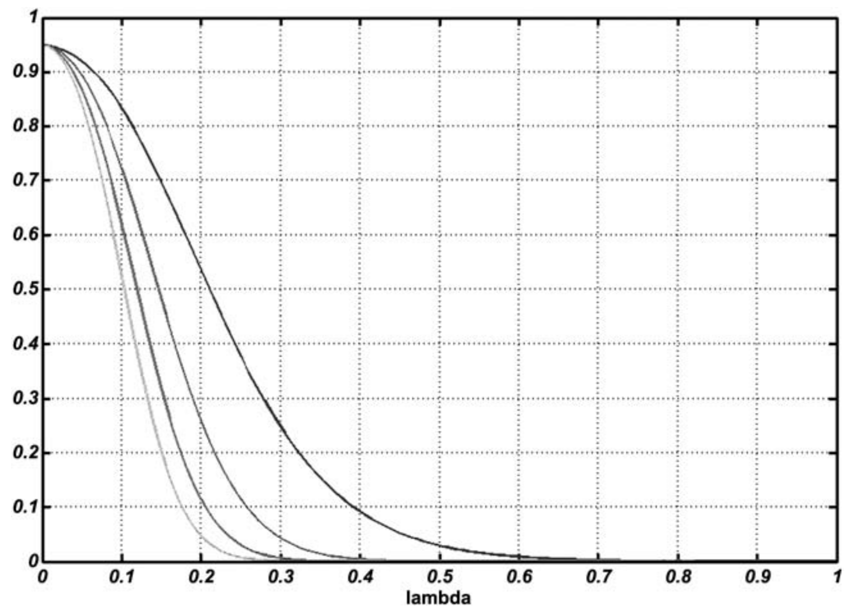

Figure 3.
Plots of θ versus λ^M with $MC_p = 1.5842$ and $n = 25(25)100$ (from bottom to top) for 95 percent confidence intervals

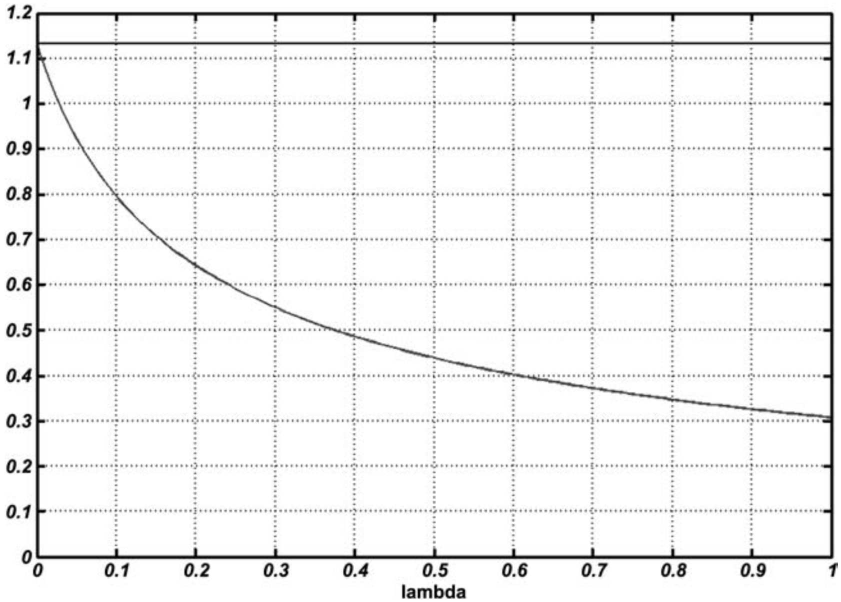

Figure 4.
Changes procedure of unadjusted lower bound versus $\lambda M = [0, 1]$

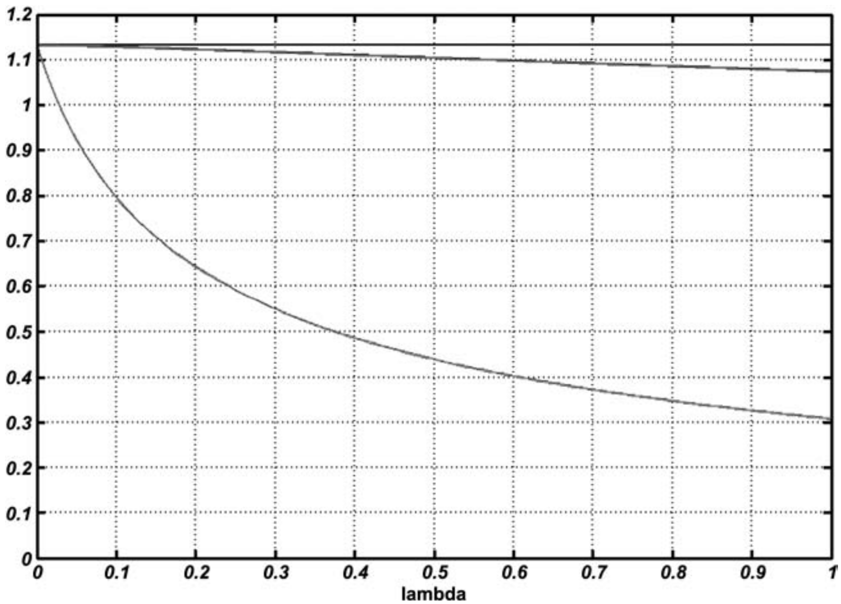

Figure 5.
Changes procedure of adjusted lower bound versus $\lambda M = [0, 1]$

from it. That is, we are trying to find out where the estimated capability is relative to either true capability, hypothesized capability, or how different the estimated and true capabilities are. To do this, we estimate an index value, compare it to a lower bound c_0, and compute the so-called p-value. The quantity p refers to the actual risk of incorrectly concluding that the process is capable of a particular test. In general, we want the p-value to be no greater than 0.05. To test whether a given process is capable, we may consider the following statistical hypothesis testing:

H0. $MC_p \leq c$ (process is not capable).

H1. $MC_p > c$ (process is capable).

where c_O is the standard minimal criteria for MC_p. The critical value, c, can be determined as:

$$P(\tilde{M}C_p \geq c_0 \big| MC_p = c) = P\left(\frac{b_v MC_p}{\sqrt{\chi^2_{n-1} \times \chi^2_{n-2} \times \ldots \times \chi^2_{n-v} / (n-1)^v}} \geq c_0 \middle| MC_p = c_O \right) = \alpha. \tag{24}$$

With respect to $y = (\chi^2_{n-1} \times \chi^2_{n-2} \times \ldots \times \chi^2_{n-v})$, then we have:

$$P\left(\frac{b_v c}{\sqrt{y/(n-1)^v}} \geq c_0 \right) = \alpha \rightarrow P\left(y \leq \frac{(n-1)^v (b_v c)^2}{c_0^2} \right) = \alpha \rightarrow F_Y^{-1}(\alpha) = \frac{(n-1)^v (b_v c)^2}{c_0^2}.$$

Thus, the critical value can be expressed as:

$$c_0 = b_v c \sqrt{\frac{(n-1)^v}{F_Y^{-1}(1-\alpha)}}, \tag{25}$$

and the power of the test (the chance of correctly judging a capable process as capable) can be computed as:

$$\pi(MC_p) = P\{\hat{M}C_p > c_0 \big| MC_p\} = P\left(y < \frac{F_y^{-1}(1-\alpha) \times (b_v MC_p)^2}{c^2} \middle| MC_p. \right) \tag{26}$$

In the presence of measurement errors, the critical value (denoted by c_0^Y) α-risk (denoted by α^Y) and the power of the test (denoted by π^Y) are:

$$c_0^Y = b_v c \sqrt{\frac{(n-1)^v}{F_Y^{-1}(1-\alpha)}} \times \frac{1}{\sqrt{(\lambda^M c)^2 + \frac{|\Sigma_Y| - |\Sigma_{\mathrm{Me}}|}{|\Sigma_Y - \Sigma_{\mathrm{Me}}|}}}, \tag{27}$$

$$\alpha^{Y} = P\left(y < \frac{(n-1)^{v}(b_{v}c)^{2}}{(c_{0}^{Y})^{2}} \times \frac{1}{(\lambda^{M}c)^{2} + \frac{|\Sigma_{Y}| - |\Sigma_{\mathrm{Me}}|}{|\Sigma_{Y} - \Sigma_{\mathrm{Me}}|}}\right), \tag{28}$$

$$\pi^{Y}(MC_{p}) = P\left(y < \frac{F_{y}^{-1}(1-\alpha) \times (b_{v}MC_{p})^{2}}{c^{2}} \times \frac{1}{(\lambda^{M}MC_{p})^{2} + \frac{|\Sigma_{Y}| - |\Sigma_{\mathrm{Me}}|}{|\Sigma_{Y} - \Sigma_{\mathrm{Me}}|}}\middle| MC_{p}\right). \tag{29}$$

With respect to equation (28); it can be seen that the right side of this probability Equation multiplied by:

$$\left[(\lambda^{M}c)^{2} + \frac{|\Sigma_{Y}| - |\Sigma_{\mathrm{Me}}|}{|\Sigma_{Y} - \Sigma_{\mathrm{Me}}|}\right]^{-1}.$$

So we underestimate the true capability of the process when we calculate process capability index using $\tilde{M}C_{p}^{Y}$ instead of $\tilde{M}C_{p}$, the probability that $\tilde{M}C_{p}^{Y}$ is greater than c_{0} will be less than the probability of that using $\tilde{M}C_{p}$. Thus, the α-risk using $\tilde{M}C_{p}^{Y}$ to estimate MC_{p} is less than the α-risk using $\tilde{M}C_{p}$ to estimate MC_{p} ($\alpha^{Y} \leq \alpha$).

Also in comparison of equation (26) with equation (29); one can see that equation (29) is multiplied by:

$$\left[(\lambda^{M}c)^{2} + \frac{|\Sigma_{Y}| - |\Sigma_{\mathrm{Me}}|}{|\Sigma_{Y} - \Sigma_{\mathrm{Me}}|}\right]^{-1}.$$

So, the power using $\tilde{M}C_{p}^{Y}$ to estimate MC_{p} is also less than the power using $\tilde{M}C_{p}$ to estimate MC_{p} ($\pi^{Y} \leq \pi$).

To improve the method of testing hypothesis for the MC_{p}, one can reconsider the testing procedure such that in the case of gauge measurement errors, a better estimation of critical region and power of the test is obtained.

If we define the $\tilde{M}C_{p}^{Y}$, using the mentioned definitions in the previous sections then:

$$\frac{MC_{p}^{Y}}{\tilde{M}C_{p}^{Y}} = \frac{b_{v}MC_{p}^{Y}}{\tilde{M}C_{p}^{Y}} \sim \sqrt{\frac{\chi_{n-1}^{2} \times \chi_{n-2}^{2} \times \ldots \times \chi_{n-v}^{2}}{(n-1)^{v}}} \Rightarrow \tilde{M}C_{p}^{Y}$$

$$= \frac{b_{v}MC_{p}^{Y}}{\sqrt{\chi_{n-1}^{2} \times \chi_{n-2}^{2} \times \ldots \times \chi_{n-v}^{2}/(n-1)^{v}}}. \tag{30}$$

Therefore the new α value is given by:

$$\alpha^{*} = P(\tilde{M}C_{p}^{Y} \geq c_{0}^{*} | \tilde{M}C_{p}^{Y} = c) = P\left(y < \frac{MC_{p}^{Y}(n-1)^{v}b_{v}^{2}}{c_{0}^{*}\left[(\lambda^{M}MC_{p})^{2} + \frac{|\Sigma_{Y}| - |\Sigma_{\mathrm{Me}}|}{|\Sigma_{Y} - \Sigma_{\mathrm{Me}}|}\right]}\middle| \tilde{M}C_{p}^{Y} = c\right).$$

Based on the above probability phrase, the new critical value is obtained as (31):

$$c_0^* = \frac{b_v \times c \times (n-1)^{v/2}}{\left[F_Y^{-1}(1-\alpha)\right]^{1/2} \sqrt{(\lambda^M c)^2 + \frac{|\Sigma_Y| - |\Sigma_{\text{Me}}|}{|\Sigma_Y - \Sigma_{\text{Me}}|}}}. \tag{31}$$

Also to improve the power function of the mentioned testing hypothesis; one can use the $\tilde{M}C_p^Y$ based on (30):

$$\pi^*(MC_p) = P(\tilde{M}C_p^Y \geq c_0^* \big| \tilde{M}C_p^Y = c)$$

$$= P\left(y < \frac{F_y^{-1}(1-\alpha) \times (MC_p)^2}{c^2} \times \frac{(\lambda^M c)^2 + \frac{|\Sigma_Y| - |\Sigma_{\text{Me}}|}{|\Sigma_Y - \Sigma_{\text{Me}}|}}{(\lambda^M MC_p)^2 + \frac{|\Sigma_Y| - |\Sigma_{\text{Me}}|}{|\Sigma_Y - \Sigma_{\text{Me}}|}} \right) (MC_p). \tag{32}$$

For the Chen example, assume that we want to test the following hypothesis:

H0. $MC_p \leq 2$.

H1. MCp > 2.

Using (27) and (28), the critical value (c_0^Y), and alpha value (α^Y) are calculated 3.0179 and 0.001 consequently; so the calculated α^Y is considerably smaller than the significant level of the test (($\alpha Y = 0.001) < \alpha = 0.05$)), and it will cause to accept the null hypothesis in many cases.

Accepting the null hypothesis means rejecting the actual capability of the process with respect to consumer view, therefore it is essential to use the measurement error for testing hypothesis using (31) in order to avoid the false decision.

For the Chen (1994) example the critical value using (31) is $c_0^* = 2.1102$. Using this value for testing hypothesis we get the desired α value ($\alpha^Y = 0.005$). Now if the capability index is increased to $MC_p = 3$ then the power of the test without considering the measurement error (29) is given as $\pi^Y(MC_p) = 0.0677$. If the measurement error is taken in to account for evaluating the power of the test (32) then $\pi^*(MC_p) = 0.5521$. Comparing these two values show that taking in to account the gauge measurement errors will cause a great deal of improvement in testing hypothesis for process capability index.

According to Figure 6, one can see that the increase in measurement capability index and sample size will considerably decrease the value of α_M^Y which will cause the under-estimation of process capability index and may conclude the acceptance of the null hypothesis that the process is not capable of doing its job.

Recalculating the critical value using (31), we get a more accurate estimate of MC_p and the critical value (c_0^*) with the exact significant level of 5 percent, which will improve the procedure of testing hypothesis.

The effect of measurement capability index on the unadjusted power of the test (π^Y) for different sample size and $c = 2$ is shown in Figure 7. As one can see the power of the test will decrease by increasing the measurement capability index. This is not the case for adjusted power of the test (π^*) as it is shown in Figure 8, therefore by considering the measurement error and its index for testing hypothesis using (32) we

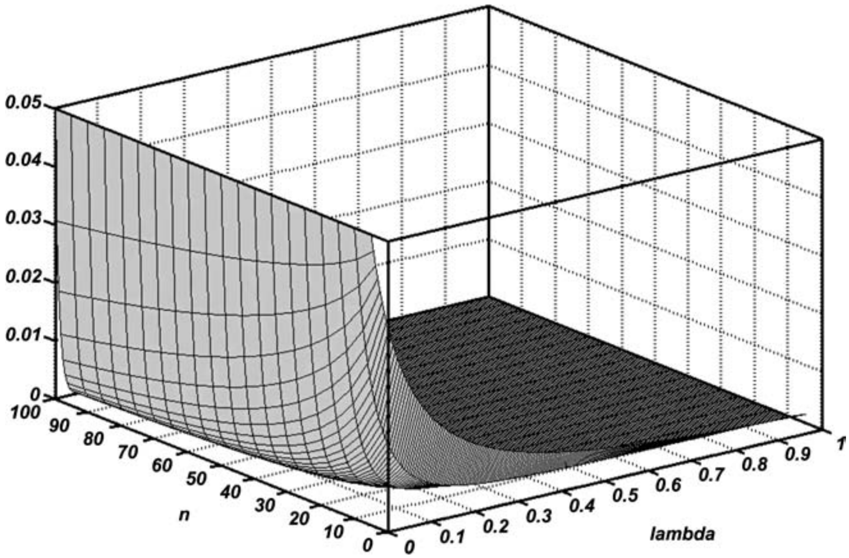

Figure 6.
Surface plot of α^Y with $n = 5(1)100$ and λ^M in $[0, 1]$ for $c = 2$ and $\alpha = 0.05$

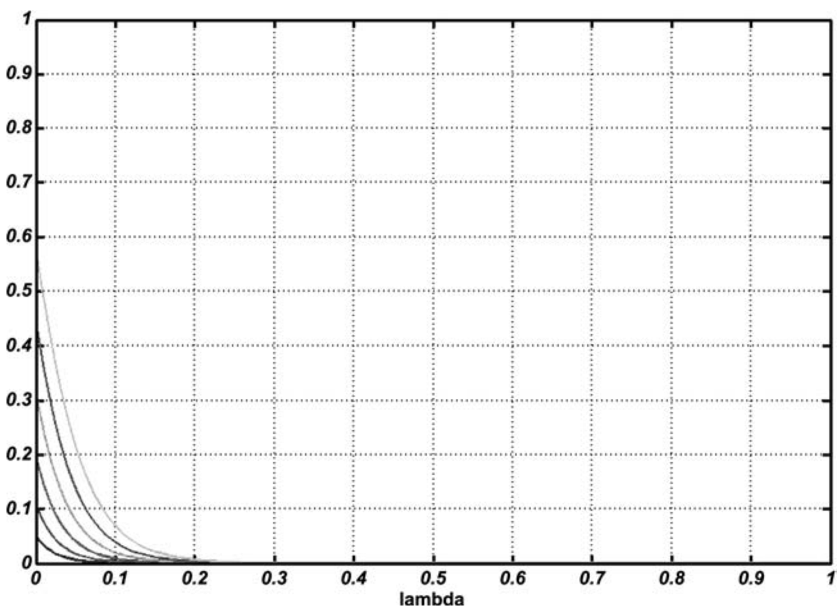

Figure 7.
Changes procedure of π^Y versus $\lambda^M = [0, 1]$ with $n = 25$ and $\alpha = 0.05$ for $c = 2$ and $MC_p = 2.00(0.20)3.00$ (from bottom to top)

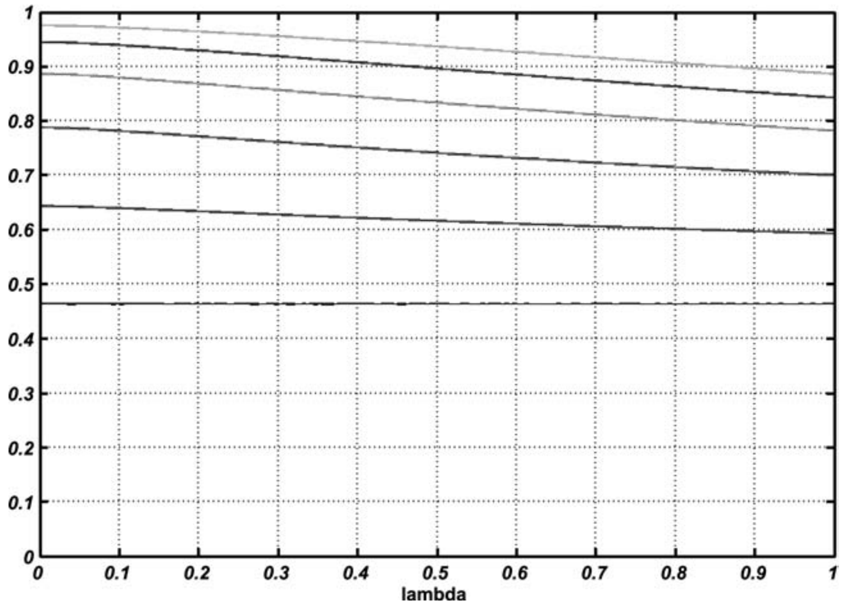

Figure 8.
Changes procedure of π^* $\lambda^M = [0, 1]$ with $n = 25$ and $\alpha = 0.05$ for $c = 2$ and $MC_p = 2.00(0.20)3.00$ (from bottom to top)

get a more accurate procedure for testing hypothesis about the capability index of the process in multivariate case.

Conclusions

With respect to the results obtained in this paper, it is specified that gauge measurement capability has an important effect on determining the process capability and this effect grows with increasing the gauge measurement error. It is also emphasized that the quality of data on the process characteristics relies very much on the gauge accuracy. Any variation in the measurement process has a direct impact on the ability to make sound judgment about the manufacturing process. Conclusions about capability of the process without considering the gauge measurement capability are not reliable and in general, producers will be damaged if they ignore this fact. We also claim that by considering the gauge measurement capability for estimating the process capability, we get a better estimation. Consequently, the more real estimations of process capability help the quality experts to adopt a better condition to improve the quality of the production.

Also we showed that the confidence coefficients may become insignificant, and that the α-risk and the power of the test may decrease with a significant magnitude due to gauge measurement errors, resulting with too understating capability. It could be a serious loss to the producers if gauge capability is not considered in process capability estimation and testing. Since measurement errors may not be avoided, having proper confidence coefficients and power becomes essential. Thus we provided adjusted confidence bounds and critical values for practitioners to use in determining whether their processes meet the capability requirement. For future researches studying other multivariate process capability indices with considering the gauge measurement capability was suggested.

References

Boyles, R.A. (1996), “Exploratory capability analysis”, *Journal of Quality Technology*, Vol. 28 No. 1, pp. 91-8.

Chan, L.K., Cheng, S.W. and Spiring, F.A. (1991), “A multivariate measure of process capability”, *Journal of Modeling and Simulation*, Vol. 11, pp. 1-6.

Chen, H. (1994), “A multivariate process capability index over a rectangular solid tolerance zone”, *Statistica Sinica*, Vol. 4, pp. 749-58.

Chen, J.P. and Cherng, G.D. (2001), “A new process capability index for non-normal distributions”, *International Journal of Quality & Reliability Management*, Vol. 18 No. 7, pp. 762-70.

Karl, D.P., Morisette, J. and Taam, W. (1994), “Some applications of a multivariate capability index in geometric dimensioning and tolerancing”, *Quality Engineering*, Vol. 6, pp. 649-65.

Kotz, S. and Johnson, N.L. (2002), “Process capability indices – a review: 1992-2000”, *Journal of Quality Technology*, Vol. 34 No. 1, pp. 1-19.

Levinson, W.A. (1995), “How good is your gauge?”, *Semiconductor International*, October, pp. 165-8.

Mittag, H.J. (1997), “Measurement error effects on the performance of process capability indices”, *Frontiers in Statistical Quality Control*, Vol. 5, pp. 195-206.

Montgomery, D.C. (1996), *Introduction to Statistical Quality Control*, 3rd ed., John Wiley & Sons, NewYork, NY.

Montgomery, D.C. and Runger, G.C. (1993), “Gauge capability and designed experiments, Part I: basic methods”, *Quality Engineering*, Vol. 6 No. 1, pp. 115-35.

Pearn, W.L. and Kotz, S. (2006), *Encyclopedia and Handbook of Process Capability Indices*, Series on Quality, Reliability and Engineering Statistics, Vol. 12, World Scientific Publishing, Singapore.

Pearn, W.L., Kotz, S. and Johnson, N.L. (1992), “Distributional and inferential properties of process capability indices”, *Journal of Quality Technology*, Vol. 24 No. 4, pp. 216-31.

Pearn, W.L., Wang, F.K. and Chen, J. (2007), “Multivariate capability indices: distributional and inferential properties”, *Journal of Applied Statistics*, Vol. 34 No. 8, pp. 941-62.

Shahriari, H., Hubele, N.F. and Lawrence, F.P. (1995), “A multivariate process capability vector”, *Proceedings of the 4th Industrial Engineering Research Conference, Institute of Industrial Engineers, Nashville, TN, May 24-25*, pp. 304-9.

Shishebori, D. and Hamadani, A.Z. (2008), “The effect of gauge measurement errors on multivariate process capability”, *Proceedings of the 3rd World Conference on Production and Operations Management (POM) Tokyo, 5-8 August*, Chapter 17, pp. 24-32.

Sultan, T. (1986), “An acceptance chart for raw material of two correlated properties”, *Quality Assurance*, Vol. 12, pp. 70-2.

Taam, W., Subbaiah, P. and Liddy, J.W. (1993), “A note on multivariate process capability indices”, *Journal of Applied Statistics*, Vol. 20 No. 3, pp. 339-51.

Wang, F.K. and Chen, J.C. (1998), “Capability index using principal component analysis”, *Quality Engineering*, Vol. 11 No. 1, pp. 21-7.

Wang, F.K. and Du, T.C.T. (2000), "Using principal component analysis in process performance for multivariate data", *OMEGA: the International Journal of Management Science*, Vol. 28 No. 2, pp. 185-94.

Wang, F.K., Miskulin, J.D. and Shahriari, H. (2000), "Comparison of three multivariate process capability indices", *Journal of Quality Technology*, Vol. 32 No. 3, pp. 263-75.

Further reading

Castagliola, P. and Castellanos, J.V.G. (2008), "Process capability indices dedicated to bivariate non-normal distributions", *Journal of Quality in Maintenance Engineering*, Vol. 14 No. 1, pp. 87-101.

Jagadeesh, R. and Subash Babu, A. (1994), "Process capability assessment with tool wear: an investigative study", *International Journal of Quality & Reliability Management*, Vol. 11 No. 2, pp. 51-62.

Pearn, W.L. and Liao, M.Y. (2007), "Estimating and testing process precision with presence of gauge measurement errors", *Quality and Quantity*, Vol. 41 No. 5, pp. 757-77.

Corresponding author

Ali Zeinal Hamadani can be contacted at: hamadani@cc.iut.ac.ir

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0265-671X.htm

Multi-response robust screening in quality construction blue-printing

George J. Besseris
Technological and Educational Institute of Piraeus, Piraeus, Greece, and The University of the West of Scotland, Paisley, UK

Received 13 November 2008
Revised 2 December 2008
Accepted 26 January 2009

Abstract

Purpose – The aim of this paper is to circumvent the multi-distribution effects and small sample constraints that may arise in unreplicated-saturated fractional factorial designs during construction blueprint screening.

Design/methodology/approach – A simple additive ranking scheme is devised based on converting the responses of interest to rank variables regardless of the nature of each response and the optimization direction that may be issued for each of them. Collapsing all ranked responses to a single rank response, appropriately referred to as "Super-Ranking", allows simultaneous optimization for all factor settings considered.

Research limitations/implications – The Super-Rank response is treated by Wilcoxon's rank sum test or Mann-Whitney's test, aiming to establish possible factor-setting differences by exploring their statistical significance. An optimal value for each response is predicted.

Practical implications – It is stressed, by example, that the model may handle simultaneously any number of quality characteristics. A case study based on a real geotechnical engineering project is used to illustrate how this method may be applied for optimizing simultaneously three quality characteristics that belong to each of the three possible cases, i.e. "nominal-is-best", "larger-is-better", and "smaller-is-better" respectively. For this reason, a screening set of experiments is performed on a professional CAD/CAE software package making use of an $L_8(2^7)$ orthogonal array where all seven factor columns are saturated by group excavation controls.

Originality/value – The statistical nature of this method is discussed in comparison with results produced by the desirability method for the case of exhausted degrees of freedom for the error. The case study itself is a unique paradigm from the area of construction operations management.

Keywords Process planning, Design and development, Construction industry, Geotechnical engineering, Technical drawing

Paper type Research paper

Introduction

Optimization of multiple quality characteristics under a host of design factors has been proven to be a subject of paramount importance in product and process development. This is understandable because it is a rare occasion, under normal conditions, that a single response would suffice to describe the trait spectrum of a product or a process. As a result, the multi-response optimization problem has been attacked from various angles where each time known powerful statistical and data mining techniques are utilized successfully to a specific end. Often, in real product development settings,

The author would like to thank the Editor-in-Chief, Professor Ton van der Wiele, and the reviewers for the efficient handling of this paper.

International Journal of Quality & Reliability Management
Vol. 26 No. 6, 2009
pp. 583-613

0265-671X
DOI 10.1108/02656710910966147

budget restrictions and time limitations may require minimal experimental effort. Under these constraints, an experimental scheme may necessitate the concurrent optimization of several quality control factors against a group of characteristics. From a technical perspective, seldom in literature one finds cases where all three possible optimization directions, "larger-is-better", "smaller-is-better" and "nominal-is-best" have been dealt with at the same time.

The level of importance of screening experiments has been elevated dramatically in the past decade as multifaceted need for technological innovation permeates traditional areas in manufacturing, logistics, medical services, telecommunications and other complex environments. Misjudgment during screening has detrimental effect on parameter design that follows. Errors due to subjective factor elimination during screening selection will be hard to be uncovered later on. When such errors gone undetected, unexplained variance may be strengthened and hence a loss of factor importance during parameter design lurks. Indeed this is double trouble. Loss of information due to screening will induce further loss of information by the time parameter design is completed. Occasionally, experienced practitioners may recover information lost during screening while in the parameter design phase and make appropriate corrections. By then there is mounting cost and loss in time and credibility that goes with it. In this excessively competitive age, all this may end up to a final product that missed the chance to get to the market first. Thus, screening may also appear as the main driver in a design work. Looking at it in a more familiar perspective, modern quality philosophy preaches to "do it right from the start" (Kackar, 1985). In DOE, screening is the start. However, in this work, it will be given emphasis on robust analysis methods especially when unreplicated designs are enforced because of economic or overall project time limitations. The method presented here may be considered as the multiresponse analog of an approach to the existing techniques for the problem of unreplicated single response optimization in fractional factorial designs.

This paper attempts to solve simultaneously all three possible optimization directions leading to a set of optimal factor settings in statistical terms for unreplicated saturated fractional factorial designs with zero degrees of freedom and zero error for all implicating factors. The technique may be useful for either screening problems or ordinary parameter design optimization alike. In case it is used for ordinary parameter design the linearity of factors against the responses has to be secured unless the working range the factor is examined is narrow enough to approximate a linear behaviour. The difficulty encountered solving unreplicated saturated fractional factorial designs for a single response has been well documented (Box *et al.*, 2005; Wu and Hamada, 2000). This difficulty stems from the fact that almost all techniques have to introduce some kind of subjective rule to overcome estimation complexities and uncertainties. It is understood that this problem would be greatly magnified when the problem is upgraded to the multi-response optimization case. However, to the knowledge of this author such discussion has not been directed to the multiple characteristics optimal control when an attempt is made to screen out factors that may not deserve to become key players later on in the parameter design phase while still maintaining the saturated fractional factorial design status. The proposed solution is attained by introducing simple blending rules based on individual response ranking methods for optimization in a suitable fractional factorial design. The statistical significance of each factor is obtained by applying the rank-sum method of Wilcoxon (1945), and Mann and Whitney (1947) to a single consolidated and re-ranked response column. The experimental arrangement of factors and levels is exemplified by capping the proposed method with an application that investigates seven possible factors in an $L_8(2^7)$ orthogonal array. However, the technique presented here will be suitable for cases where orthogonal arrays of the family: $L_4(2^3)$, $L_8(2^7)$, $L_{12}(2^{11})$ and $L_{16}(2^{15})$ are adopted for short-cycle and economic experimentation (Taguchi *et al.* (2000)). The advantages of employing orthogonal arrays have been extensively discussed by Dey (1985). The technique may be proved to be convenient and practical for cases where the experimental runs are restricted by budget or time to an absolutely minimum effort. To demonstrate the potential of this method a case study is drawn from the field of geotechnical engineering and more precisely from the field of construction design optimization that is based on computer experiments performed on a real excavation project. The quality responses that are defined in this example are selected to cover all three possible optimization directions. Specifically, the safety factor is sought to be maximized, the steel tension to be minimized and the total displacements to reach a target value often predetermined by local and international construction legislation. The technique presented here easily may be extended to any number of quality characteristics. The problem of weighing the importance of some responses against the others is not addressed in this work even though it is rather obvious how weighting factors may adjust the influence of individual quality characteristics to the total sum of all responses. Nevertheless, the engineers in the case study presented later in this article, had not a definite preference scheme for the responses accounted. In a forthcoming publication, the effect of predetermined weights on the studied responses with respect to the final list of active factors will be investigated. As expected, it is finally shown that concurrent optimization may not select out the factor settings that would optimize each response separately.

Literature review

The foundations of experimental design methods from a modern perspective are outlined in the standard treatises of Box *et al.* (2005) and Montgomery (2004). Industrial applications of DOE methods with concepts of the signal-to-noise ratio can be found in Taguchi *et al.* (2000), Phadke (1989) and Ross (1988). Contemporary review of the field can be found in Arvidsson and Gremyr (2008), and Robinson *et al.* (2004). Derringer and Suich (1980) attempted to extend the experimental design theory to situations where a product developer has to decide on optimal control factors for cases that involve more than one quality characteristic simultaneously by adopting the desirability function (Harrington, 1965). The single aggregate measure that captures the distance between responses and the ideal point has been studied by Khuri and Conlon (1981). Del Castillo *et al.* (1998) provided improved versions of the desirability function for multi-response optimization cases. Elsayed and Chen (1993) proposed production-driven analytical tools in order to achieve optimal levels for process parameters for products with multiple characteristics. Kim and Lin (2000) modelled a multi-response system by proposing an exponential form of the desirability function aiming at simplifying the desirability function assessment process. Ames *et al.* (1997) utilized the quality loss function to optimize multiple response surfaces. Vining (1998) showed that, for practical reasons, a compromise may produce optimal results for a case where various quality characteristics are evaluated at the same time. Reddy *et al.* (1997) attempted to unify robust design and goal programming for multi-response optimization. Antony (2000) attempted to reduce the number of effective quality characteristics from two to one by employing principal component analysis on normalized response values and then performed Analysis of Variance (ANOVA) to the resulting single column that carried the blended information from those two quality characteristics. Kumar *et al.* (2000) evaluated the quality of three features of V-processed castings of an aluminium-silicon alloy by joining Taguchi's technique with the utility concept. Taguchi's dynamic system consideration of biological reduction of ethyl acetoacetate process experiment at the Union Chemical Laboratories was analyzed by multi-regression analysis coupled with the desirability function (Hsieh *et al.*, 2005). An injection moulding application of Taguchi's method involving ANOVA and ANOM techniques were elucidated on a multiple-characteristic improvement study on washing machine agitator (Reddy *et al.*, 1998). Hsieh and Tong (2001) provided case studies when qualitative and quantitative multiple characteristics required to be optimized in IC manufacturing by applying a method of neural networks. Tong and Wang (2002) combined principal component analysis and grey relational analysis targeting multi-response optimization for problems of industrial interest. Tang and Xu (2002) proposed a comprehensive method for the simultaneous search for optima when two quality characteristics are considered by employing the RSM statistical analysis. Kros and Mastrangelo (2001) showed in two examples that the Global Criterion method results in multiple optimal points in contrast to the desirability method that results in one set for mixed types of multi-response systems. Berni and Gonnelli (2006) have employed Response Surface Methodology to optimize a multi-response accuracy measurement in a numerical control machine. Wurl and Albin (1999) investigated the sensitivity of the optimal control factors in cases with multiple quality characteristics.

Often an experimental proposal may require a short listing of trials because either each run may turn out to be expensive owing to the implicating materials and methods or simply due to the timeliness of the results if it is to be considered practical from a professional perspective. On the other hand, small sample experimentation lends itself to questions regarding the distribution of the responses before powerful techniques such as Analysis of Variance and RSM are to be employed. Averting the complication of distribution function determination can be successful by adopting a scheme based on ranking methods (David and Nagarajia, 2003). Conover (2006) provided modern insight in some specialized multi-factor contrasting using the rank transform method. Furthermore, nonparametrics are more suitable for restricted data sampling because they make use of the median that provides a robust central inference (Huber, 2003). Moreover, a simple ranking technique has been presented based on the Wilcoxon-Mann-Whitney approach applied on the eight-run orthogonal array (Besseris, 2008). Order statistics have not been exploited as of yet for the multi-response case. This will be the subject of the present report.

Recently, new methods associated with the quality improvement of designing parameters may employ experimentation directly on CAD/CAE prototypes by making use of DOE methods. In essence, DOE methods aid engineers to make meaningful selections among several possible pre-designed model candidates with the assistance of computer experiments (Kenett and Steinberg, 2006). There is extensive literature in

DOE methods along with applications drawn on a plethora of case studies from the manufacturing environments (Tong *et al.*, 1997; Jayaram and Ibrahim, 1999; Tong and Hsieh, 2000; Antony, 2001; Lu and Antony, 2002; Liao and Chen, 2002; Liao, 2003; Antony *et al.*, 2005). Quality pre-design has driven an increasing need for simulations in order to cut down cost and product validation time. To this end, computer simulated experiments have been devised and been compared among classical and more modern approaches with respect to the concept of robustness in product and process design (Bates *et al.*, 2006). However, the theoretical fundamentals of this field have been introduced earlier by Sacks and Welch (1991) and Sacks *et al.* (1989a, 1989b) where the theory of analyzing large-run computer generated data for a circuit-simulator case was outlined taking in account the simulator complexity. Romano *et al.* (2004) have provided an insight of how CAD/CAE code may be optimized in terms of intrinsic engineering and model factors. The necessity of simulations in modern day decision-making has been well-exposed (Moorthy (1998), Moorthy (1998), Purcell (1999) and Schrage (2000)).

Methodology for non-parametric multi-response optimization

The methodology outlined in this paragraph is intended mainly for industrial and other process-focused, quality-improvement actions where a small number of trials is needed because of economical reasons or severe time limitations. Adherence to normality is not necessary to carry out this procedure. The topic of weighing the investigated quality characteristics according to a desirability scheme is easily anticipated as an extension to the work provided here but it would not be undertaken explicitly in this work. The steps that should be followed to a successful determination of any number of quality responses can be enumerated in eight distinct stages. These are:

(1) *Problem determination.* At the inception of a quality improvement effort, the importance of the solution to the problem needs to be established through standard quality management tools and techniques such as a Pareto analysis or a PPM analysis. Brainstorming sessions answer the questions regarding the severity, the detectability, the locality, and the occurrence rate of the problem. Usually, a problem statement is stated at the end of this step.

(2) *Quality response selection.* It is important to identify the quality characteristics that influence process or product performance early. Along with this issue, measurement systems, methods and people that need to carry out the data collection step are addressed at this point.

(3) *Factor selection.* Controlled factors along with controlled and uncontrolled noise are selected from examining process flow diagrams, process control plans and product design specifications. At this stage, the relationship between the factors and the noise against the various responses is determined in terms of simple functionality rules such as the possible existence of a linear or quadratic type of dependence among them and so forth. Finally, the various settings, or levels, at which each factor will be tested are also identified at this time. Depending on previous experience on the physical problem under consideration, it is instructive to decide at this stage the possibility of interactions between factors

and the extent of their engagement, i.e. two-factor, three-factor interactions and so forth.

(4) *Experimental design.* Orthogonal array tables are selected depending on the number of factors, noises and interactions determined from the previous step to provide an economical recipe for trial run execution. Also, the randomization scheme, possible blocking and the repetitions necessary to provide adequate resolution between the effects to be measured in the next stage are also concluded at the experimental design phase.

(5) *Trial run execution.* Trial runs are performed according to the plan of step 4 and measurements are collected for all recorded test combinations. Verification of the suitability of the sample size used is obtained at this stage. The data sheet is filled. Missing repetitions and trials may be handled as unbalanced data sets, or be replaced by known values from other information derived from the central tendencies of the rest of the data or even rerun the missing experiments.

(6) *Data analysis.* Data analysis tools are utilized to uncover the optimum combinations of statistically influential factors and noise. A predictive value of the quality characteristics is estimated according to the optimum settings resulting from the analysis. Usually, this step is carried out on statistical software where central tendencies, variations and graphic comparisons are contrasted to optimize the quality characteristics. Specifically, for this work non-parametric theory is accommodated to fractional factorial designs to support the multi-response optimization and it is discussed separately in the next chapter.

(7) *Confirmatory test execution.* The experiment is repeated several times to confirm the optimal settings of factors and the level of the noise present. What is expected at this stage is the validation of the prediction values of the quality characteristics calculated in the previous step. The statistical significance of these optimal settings along with the generated predictions is finalized in this step in terms of confidence intervals.

(8) *Conclusions.* The efficacy of the experimental work completed and its importance in furthering the quality improvement effort in the selected process or product characteristics is summarized in the final step.

Multi-response super-ranking formulation in fractional factorial design

Montgomery (2004), Wu and Hamada (2000), Phadke (1989) and Ross (1988) have proposed that fractional factorial designs are suitable for experimentation when there is a relatively large number of influencing factors that need to be investigated comparing to the total allowed number of trials. Let us suppose that there are a total of k factors and interactions that need to be considered. The factors and interactions are given the symbols X_1, $X_2 \ldots X_k$. Let us suppose that this choice of factors and interactions calls for n experimental trials. Then, the required levels of this arbitrary orthogonal array may form columns of settings that may be represented as X_{1j}, $X_{2j} \ldots X_{kj}$ ($j = 1, 2 \ldots n$). Furthermore, we assume that there is a count of m distinct, independent responses that need to be optimized simultaneously. Each of the responses may be symbolized by Y_1, $Y_2 \ldots Y_m$, where the dimension of each response is n. Each of the responses may be visualized as vectors with elements $y_{1j}, y_{2j}...y_{mj}$ ($j = 1, 2...n$). The generic fractional factorial design is portrayed appropriately in Figure 1.

In a very simple manner, the m responses are cast in m ranked response vector fields where the elements of each response are ranked individually and independently from the elements of the other responses. Thus, the minimum rank that is given to an element in any of the responses will be one and the maximum will be n. The newly transformed variables become $yij \rightarrow Tij$. Subsequently, the ranked elements for each response are squared. This process is shown schematically in Figure 2.

At this point, it is desirable to form a measure for the combined effect of all responses at the same time. To do this, it is proposed to add all elements of the newly transformed rank-squared responses for the same row or vector position according to the scheme depicted in Figure 2.

Each summation produces a single quantity, T_j^{TOT} ($j = 1, 2...$n) that represents a single element which in turn belongs in a single "master column" of dimension n. This "master-column" receives one more rank ordering to result in the form T_j' ($j = 1, 2...n$). All that is needed at this point is to cast the multi-column responses that obviously may foster a different unit convention for each response to a single unit-less column that carries all information regarding the combined effect of a specific experimental run to all responses for all scheduled trials. Let us call this complete multi-processing of multi-response elements as "super-ranking" to distinguish it from simple rank ordering that a single response may undergo (Figure 3). It becomes

$$\begin{array}{c} \textit{FACTORS \& INTERACTIONS} \qquad\qquad \textit{RESPONSES} \\ \begin{array}{c|cccc|c} \textit{RUN} & X_1 & X_2 & & X_k & \\ 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 2 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ n & X_{1n} & X_{2n} & \cdots & X_{kn} \end{array} \begin{array}{c} Y_1 \\ \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} \end{array} \begin{array}{c} Y_2 \\ \begin{bmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{bmatrix} \end{array} \cdots \begin{array}{c} Y_m \\ \begin{bmatrix} y_{m1} \\ y_{m2} \\ \vdots \\ y_{mn} \end{bmatrix} \end{array} \end{array}$$

Figure 1. Arrangement of an orthogonal array of factors and interaction with displayed levels and their respective responses

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} \begin{bmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{bmatrix} \cdots \begin{bmatrix} y_{m1} \\ y_{m2} \\ \vdots \\ y_{mn} \end{bmatrix} \rightarrow \begin{bmatrix} T_{11} \\ T_{12} \\ \vdots \\ T_{1n} \end{bmatrix} \begin{bmatrix} T_{21} \\ T_{22} \\ \vdots \\ T_{2n} \end{bmatrix} \cdots \begin{bmatrix} T_{m1} \\ T_{m2} \\ \vdots \\ T_{mn} \end{bmatrix} \rightarrow \begin{bmatrix} T_{11}^{\ 2} \\ T_{12}^{\ 2} \\ \vdots \\ T_{1n}^{\ 2} \end{bmatrix} \begin{bmatrix} T_{21}^{\ 2} \\ T_{22}^{\ 2} \\ \vdots \\ T_{2n}^{\ 2} \end{bmatrix} \cdots \begin{bmatrix} T_{m1}^{\ 2} \\ T_{m2}^{\ 2} \\ \vdots \\ T_{mn}^{\ 2} \end{bmatrix}$$

Figure 2. Rank transformation of each response element to rank squared

$$\begin{bmatrix} T_{11}^{\ 2} \\ T_{12}^{\ 2} \\ \vdots \\ T_{1n}^{\ 2} \end{bmatrix} \begin{bmatrix} T_{21}^{\ 2} \\ T_{22}^{\ 2} \\ \vdots \\ T_{2n}^{\ 2} \end{bmatrix} \cdots \begin{bmatrix} T_{m1}^{\ 2} \\ T_{m2}^{\ 2} \\ \vdots \\ T_{mn}^{\ 2} \end{bmatrix} \xrightarrow{\textit{ElementSummation}} \begin{bmatrix} \sum_{i=1}^{m} T_{i1}^{\ 2} \\ \sum_{i=1}^{m} T_{i2}^{\ 2} \\ \vdots \\ \sum_{i=1}^{m} T_{in}^{\ 2} \end{bmatrix} \rightarrow \begin{bmatrix} T_1^{TOT} \\ T_2^{TOT} \\ \vdots \\ T_n^{TOT} \end{bmatrix} \xrightarrow{\textit{RankOrdering}} \begin{bmatrix} T_1' \\ T_2' \\ \vdots \\ T_n' \end{bmatrix}$$

Figure 3. Rank consolidation for all response elements to super-ranked form

apparent here that in this "super-ranked" form there is no need that the data distribution be investigated and identified. This resolves conveniently cases where the quality characteristics may belong to distinctly different distribution families. It should be noted at this point that squared ranks were preferred eventually. They were embodied in the method above because during exploratory work squaring proved convenient in resolving more efficiently possible "ties" that might have risen from the summation of otherwise merely non-squared ranked responses. This work does not resolve the handling and the respective effect of repetitions per specific response even though such an extension might not pose insurmountable concern. This will be the subject of a follow-up work. Finally, the fact that there might exist an indicated preference over the effect incurred on specific quality characteristics over the rest of the group may require knowledge of weighing factors for each response that it is seemingly simple to be handled by this method but again it will be set aside for investigation in an up-coming report.

In a summary, the generalized orthogonal array along with the m responses we started working with now is reduced to analyzing this array with a single "super-ranked" column that now stands as shown in Figure 4.

The next step in the analysis is to statistically differentiate the effects of each factor and/or interactions against this transformed response column, *T'*. To achieve this, the super-ranked values are grouped according to the levels dictated by each factor or interaction separately. In case the factors and interactions are tested against two settings then the Wilcoxon's Rank-Sum test or the Mann-Whitney test is the most appropriate to resolve the statistical significance. It is emphasized that a method such as the one that is proposed here should be useful for circumstances where sample count should be kept small because of economical reasons and time limitations. The strategy here is to contrast the factor setting rank sum values, for each factor separately, against each other without to take in account the influence that the other factors may have on the collected observations. If there is a need for capturing possible interactions among engaging factors then, as it is well-known, an appropriate assignment of interaction columns directly to the orthogonal array will suffice in providing resolution of such anticipated effects. The two major assumptions that also need to be stated from the beginning are:

(1) the randomness of the data collection process; and

(2) the independence of drawing these measurements.

The multi-response optimization problem in design of experiments has been outlined above in general terms. Let us refine this method further for a popular fractional factorial design such as the $L_8(2^7)$. Moreover, of great interest would be to test a case where all three possible optimization directions are witnessed simultaneously. In quality improvement efforts, these three types that arise from Taguchi's quality loss

$$\begin{array}{c} \textit{FACTORS \& INTERACTIONS} \\ \begin{array}{c|cccc} \textit{RUN} & X_1 & X_2 & \cdots & X_k \\ 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 2 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ n & X_{1n} & X_{2n} & \cdots & X_{kn} \end{array} \end{array} \begin{array}{c} \textit{RESPONSES} \\ \begin{array}{c} Y_1 \\ \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} \end{array} \begin{array}{c} Y_2 \\ \begin{bmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{bmatrix} \end{array} \cdots \begin{array}{c} Y_m \\ \begin{bmatrix} y_{m1} \\ y_{m2} \\ \vdots \\ y_{mn} \end{bmatrix} \end{array} \end{array} \xrightarrow{\textit{SuperRanked}} \begin{array}{c} \textit{FACTORS \& INTERACTIONS} \\ \begin{array}{cccc} X_1 & X_2 & \cdots & X_k \\ X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{array} \end{array} \begin{array}{c} \textit{RESPONSE} \\ T' \\ \begin{bmatrix} T_1' \\ T_2' \\ \vdots \\ T_n' \end{bmatrix} \end{array}$$

Figure 4.

concept have come to be known as optimization conditions “larger-is-better”, “smaller-is-better” and “nominal-is-best” depending on the characteristic maximization or minimization or the minimization of the response elements difference from a specified target value respectively. Thus, for the simplest case, it is required to consider at least three responses. In this fashion, the first response, denoted by Y_1, is assigned randomly to the “larger-is-better” characteristic, the second response is assigned to the “smaller-is-better” characteristic, denoted by Y_2 and the third is the “nominal-is-best” characteristic, denoted by Y_3. Another convention that needs to be specified before performing the ranking operations is that the smallest rank, i.e. rank 1, is given to:

- the response element that possesses the smallest value for the “smaller-is-better” characteristic type;
- the largest value for the “larger-is-better” characteristic type; and
- the smallest discrepancy between target value and the response element for the “nominal-is-best” characteristic type.

This convention is necessary to provide blending rules that are consistent with the total direction of the optimization. According to this notion, the smaller rank elements of a super-ranked response column point towards the concurrent optimization direction. To become even more specific, an $L_8(2^7)$ representation of triple-response set-up has the form as depicted in Figure 5.

In a two-level orthogonal array, there are two outcomes that may appear during the analysis of the data of the two settings of a factor. Either the group of values forming the data pool for one setting belongs in the same family as the other one or the two groups belong in different families. The details of the distribution(s) in either case may not be known to us in our formulation and it is not needed in order to proceed with our proposed analysis anyway. Furthermore, for practical purposes of this work, setting sample count is expected to be found in the vicinity of size four at most. It is anticipated that if it is to attempt a median value comparison between either two samples from the same distribution, or two samples from rather separate distributions, it is advisable to resort to the usage of exact tests. Let us analyze the two cases separately.

A common population for all measurements

In this case we define $X_1, X_2, \ldots, X_k$ the factors that need to be examined and X_1^1 and X_1^2, X_2^1 and $X_2^2, \ldots,$ X_k^1 and X_k^2, their respective medians of their setting outcomes which correspond to levels 1 and 2 at the L_8 orthogonal array presented above. Similarly, k is equal to seven for the same illustration. If N is the number of observations per factor and level then the observations for factor set at level 1, X_1^1, will be $X_{1,1}^1, X_{1,2}^1 \ldots X_{1,N}^1$ and

$$
\begin{array}{c}
\text{FACTORS \& INTERACTIONS} \\
\begin{pmatrix}
RUN & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\
3 & 1 & 2 & 2 & 1 & 1 & 2 & 2 \\
4 & 1 & 2 & 2 & 2 & 2 & 1 & 1 \\
5 & 2 & 1 & 2 & 1 & 2 & 1 & 2 \\
6 & 2 & 1 & 2 & 2 & 1 & 2 & 1 \\
7 & 2 & 2 & 1 & 1 & 2 & 2 & 1 \\
8 & 2 & 2 & 1 & 2 & 1 & 1 & 2
\end{pmatrix}
\end{array}
\begin{array}{c}
\text{RESPONSES} \\
\begin{array}{ccc} Y_1 & Y_2 & Y_3 \end{array} \\
\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \\ y_{18} \end{bmatrix}
\begin{bmatrix} y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{26} \\ y_{27} \\ y_{28} \end{bmatrix}
\begin{bmatrix} y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{36} \\ y_{37} \\ y_{38} \end{bmatrix}
\end{array}
\xrightarrow{\text{SUPER-RANKING}}
\begin{array}{c}
\text{FACTORS \& INTERACTIONS} \\
\begin{pmatrix}
RUN & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\
3 & 1 & 2 & 2 & 1 & 1 & 2 & 2 \\
4 & 1 & 2 & 2 & 2 & 2 & 1 & 1 \\
5 & 2 & 1 & 2 & 1 & 2 & 1 & 2 \\
6 & 2 & 1 & 2 & 2 & 1 & 2 & 1 \\
7 & 2 & 2 & 1 & 1 & 2 & 2 & 1 \\
8 & 2 & 2 & 1 & 2 & 1 & 1 & 2
\end{pmatrix}
\end{array}
\begin{array}{c}
\text{RESPONSE} \\
\begin{bmatrix} T' \\ T_1' \\ T_2' \\ T_3' \\ T_4' \\ T_5' \\ T_6' \\ T_7' \\ T_8' \end{bmatrix}
\end{array}
$$

Figure 5.

for level 2, $X^2_{1,1}$, $X^2_{1,2}$,… $X^2_{1,N}$, respectively. For the L_8 case, N will be equal to four. On this notation, the rest of the factors may also be discerned in their observations and their medians per setting. Then for each factor the null hypothesis, *Ho* and its alternative H_1, will be:

Ho: Median X^1_1 = Median X^2_1 OR H_1: Median $X^1_1 \neq$ Median X^2_1

Ho: Median X^1_2 = Median X^2_2 OR H_1: Median $X^1_2 \neq$ Median X^2_2

Ho: Median X^1_k = Median X^2_k OR H_1: Median $X^1_k \neq$ Median X^2_k where k is equal to seven for the illustration of an L_8 array.

For each factor then the powerful Wilcoxon's rank sum test is to be applied separately (Wilcoxon, 1945). The statistic that needs to be computed is the smaller sum of ranks when all observations from each setting separately are pooled into one new set while keeping the tag on for each sum of ranks for each setting. There are tables available where this statistic may be used to compare minimum critical value. If this critical value is exceeded then the null hypothesis is accepted. Otherwise, the alternative hypothesis is accepted. This concept may be extended to three-level settings per factor where in this case the Kruskal-Wallis test should be the more appropriate.

The factor-setting observations belong to separate population

Following the same thinking as in the previous case, but not conceding that all observations may belong in the same family, an alternative strategy of handling this situation is offered by altering the hypothesis testing to include the possibility that each setting' s median is comprised from a separate distribution. Again the distributions of the populations for each level per factor need not to be known. In this case the Mann-Whitney (Wilcoxon's paired test) is employed to resolve possible statistical significance of response data between the settings of a factor (Mann and Whitney, 1947). For cases where the sample is larger than four for each factor setting, the test statistics for the Wilcoxon' rank sum test, the Mann-Whitney test follow a distribution accommodating any data population, but this is not the subject of this paper. More discussion on this may be found in Conover (2006).

For illustrational purposes the responses from the super-ranked *T'* response can be grouped according to X_1 factor as $T_{L1} = T_1' + T_2' + T_3' + T_4'$ and $T_{L2} = T_5' + T_6' + T_7' + T_8'$, where T_{L1} and T_{L2} are the summed super-ranks for level 1 and level 2 of the X_1 factor of the orthogonal array L_8, respectively. The smaller of the two super-ranked sums provides the so-called Wilcoxon's "*T*-statistic" that is to be compared to the lower value of the critical interval for a specified significance level (Wilcoxon, 1945). If this sum-rank is found to be less than the lower interval limit then the null hypothesis does not hold. By the same token, the same discussion applies when the larger of the two sum-ranks is used in comparison with the high end of the critical interval. If the sum-rank exceeds this maximum value again the null hypothesis is rejected. In summary, a method was developed here that may handle the multi-response optimization screening problem for two-level unreplicated, saturated fractional factorial designs considering the difficult case where there is untraceable unexplainable error for each investigated response along with complete exhaustion of all degrees of freedom for the error. As it was stated earlier, this rather naïve method intends to compete for offering a statistical solution to the multi-characteristic optimization problem in practical situations when sampling is restricted to a few runs. Such restrictions are common when the cost of a single run is large or the time for producing results too short for a full-fledged investigation. Sampling schemes that may be benefited by this analysis are OA-type of experimental saturated designs targeting mainly the family members: $L_4(2^3)$, $L_8(2^7)$, $L_{12}(2^{11})$ and $L_{16}(2^{15})$ (Taguchi *et al.*, 2000).

On the Squared-Sum-of-Ranks metric

The method presented in this work relies heavily on a two-phase rank transformation. In the first phase, a simple conversion of response observables to rank variables for a given m-count of characteristics is attained. This way all responses are homogenized and they can be conveniently represented in a vector space in a proposed coordinate system. This makes the representation of data points more easily to understand and handle. In the second phase, ranking is utilized to treat the total sum of squared ranks generated from the previous phase across all considered responses for each trial run separately. This action forces a much-needed relationship between the total effect produced by all responses owing to the conditions experienced in a specific experimental run. What it may not be immediately realized from the presentation in the preceding sections is the rationale that leads from one rank transformation to the next. Therefore, a more detailed explanation is offered here. Right after it was succeeded to unify all original data observables under a single measure scale, that of ranks, it becomes apparent that the next thing is to find a way to combine the information generated by each and every implicated response to a single metric for each run separately. Only when such a quantity has been conceived, then it makes sense to be introduced to a robust statistical processor (such as the Wilcoxon-Mann-Whitney test) in order to proceed with a comparative analysis. This way dependency of this anticipated "super" metric on the controlling factors may be explained in a statistical fashion. It is easily seen that what it should be attempted here is to create a global metric among all considered responses that describes judiciously in a very simple manner the outcome of an individual run. If it is to be an objective measure for the purpose it is intended to serve, it is preferable this metric to agree with widely accepted concepts in measure theory while at same time be as free of any empiricities as possible (Taper and Lele, 2004). One such measure may be the all-familiar m-dimensional Euclidean norm suitably defined and well understood in statistical theory as well as in mathematical analysis (Hazod and Siebert, 2001). By observing in Figure 3 the T_j^{TOT} ($j = 1, 2\ldots n$) quantity, we figure that this is nothing more than the squared value of the m-dimensional Euclidean norm (SEN) representing the total magnitude effected in a single j experimental run for all random and independent rank states that reflect all m-responses simultaneously. Strictly speaking, the T_j^{TOT} ($j = 1, 2\ldots n$) "super" metric does not actually need to be subsequently operated by the square root in order to have the formal norm meaning. This is because T_j^{TOT} ($j = 1, 2\ldots n$) accepts only integer values equal or greater than one. Therefore, inequalities maintained between squared values of the norm are automatically hold after taking the square root on the values of T_j^{TOT} ($j = 1, 2\ldots n$). Concluding, rank order inequalities among T_j^{TOT} ($j = 1, 2\ldots n$) elements are maintained in the same direction no matter if the these elements are operated with the square root or not. This last ordering among T_j^{TOT} ($j = 1, 2\ldots n$) values creates the Super Rank independent and random variable previously referred to as Super Rank (SR) vector.

On the test performance

For a given OA, the performance of the test is independent of the count of responses involved. This is because the range of the SR values is dictated by the run count of the OA, not by the count of the responses. For example, let's say we consider an $L_8(2^7)$ OA, then Wilcoxon-Mann-Whitney test will process four entries for each factor setting irrespectively of the count of the responses examined. Therefore, as soon as the SR column has been formed the test performance parallels actually that performance of the Wilcoxon-Mann-Whitney test. For sampling for a short-run scheme, the overall superiority of the Wilcoxon-Mann-Whitney test has been well-established (David and Nagarajia, 2003; Conover, 2006). This is one great advantage of the proposed method. Subjectively, it may seem that things may get out of hand because of the proposed successive additions of squared ranks as the number of responses increase. However, this should not be of a primary concern. According to the previous section, the sum of squared ranks across all responses for each run may be visualized as the squared norm of m-dimensional random elements (belonging in the positive integers of the real number system, (^{m}) thus forming a subset of the Euclidean vector space. The measurability properties of the metric defined by summing the squares of the response ranks across a run should coincide then with the measurability properties of the Euclidean norm for a given finite count of response elements. This warrants the objective comparison of squared norms among runs.

Case study: construction quality improvement during the design phase

Construction quality management is an area of great interest internationally because of the direct link that exists between the well-known leading economic indices and the construction development. Among the peculiarities that the field of construction possesses is that each project is a brand new project by itself. Thus, most of the experience gained in this field is best exploited very early in the planning phase of such project. As already discussed previously in this article, quality improvement has moved to a pre-designing stage by employing a combination of DOE methods and design simulations that in turn may feed design data to a product or process without the need of actual physical experimentation (Kenett and Steinberg, 2006). Construction quality management may be benefited by these latest developments by expecting to make key design decisions through DOE at early project phases assisted by Computer Aided Design (CAD) and Computer Aided Engineering (CAE) software packages. The case study that is outlined here concerns a real project on contract that involves a geotechnical study of temporary bracing (shoring) slope excavation project, for a six-floor building with ground stores and a three-floor underground garage. More details of the design can be found in Telis and Besseris (2007). The construction company management team, consisting of the operations manager, quality manager, a supervisor civil engineer and a design engineer, decided that they would test several implicating factor combinations by taking advantage of a fractional factorial experimental design engaging key quality characteristics. Previously, all design adjustments were based on experience only. At this attempt, the three responses that was decided that they should receive simultaneous optimization at the design phase were:

(1) the safety factor;
(2) the total displacements; and
(3) the steel tension.

In brief, the safety factor is an index that weighs the shearing parameters affected in the groundwork against the shearing parameters at failure point following the so-called Mohr-Coulomb criterion (Day (1999), Wood (2004)). Typical values for this response have been specified by national agencies around the globe to be 1.4 at minimum. However, this response is treated in this work as "larger-is-better". Total displacements or deformations, measured in millimetres, represent the expected locomotion of the bracing system and the ground masses. The accepted total displacement value for the foundation with individual pads is 50 mm at maximum. However, an "internal" specification of 10 mm was to be enforced such as to take in account flexibility issues of the structure. This response is treated as "nominal-is-best" in this work. Finally, steel tension, measured in T/cm^2, is the ratio of maximum propensity per pile which is caused by the ground impulses over the resistance propensity of every beam HEB IPB. A maximum value of 1.6 T/cm^2 was allowed for this project. Nevertheless, this response is treated in the optimization procedure as "smaller-is-better".

In a brainstorming session, that involved the aforementioned group, it was concluded that the following seven controlled factors needed to be investigated against all three quality characteristics. These factors are familiar to civil engineers and they are:

(1) the reinforcement;
(2) the pile depth;
(3) the anchor depth;
(4) the distance between piles;
(5) the angle of the excavation;
(6) the anchor stretch force; and
(7) the anchor length.

Along with the control factors in Table I, it is also listed the selected factor settings that the operations management team decided that this investigation should be restricted to. On the same table, the appropriate measurement units are also quoted. The selection to work with the two end-points of the indicated range was supported by:

(1) experience gained from past similar projects and existing bibliography on geotechnical engineering; and
(2) on preliminary graphical representation of each factor on one-to-one dependence check with each the quality characteristic separately.

	Control factor	Level 1	Level 2
A	Reinforcement	HEB IPB 120	HEB IPB 180
B	Pile's Depth (m)	− 10.50	− 12.00
C	Anchor Depth (m)	− 4.00	− 3.00
D	Pile's Distance (m)	1.00	1.50
E	Angle of the Excavation	3:4	4:3
F	Anchor Stretch Force (kN/m)	100	200
G	Anchor Length (m)	12.00	14.00

Table I.
Control factor and level settings for control factor optimization

The experiments were run on Plaxis (v.8), a professional construction design software package that is essentially a CAD/CAE finite-elements-based system intended to provide two-dimensional analysis of deformation and stability of large-scale geotechnical engineering designs. The total time to complete a program inputting, program execution and output decoding is about 20-30 minutes. Thus, by default, this case calls for a low count of experiments if the method outlined in this work is to be practical and convenient at the same time. This would definite provide a motive for introducing this quality tool as a standard work procedure to every project. Since, seven factors are considered, then an $L_8(2^7)$ is appropriate for a "screening" set of experiments. However, high repeatability these experiments naturally possess compensates for the time-consuming program execution process. Indeed, on trial checks the accuracy was found consistent to the third decimal figure for all three responses. This is in accord with the level of accuracy of modern finite-element packages. Hence, no repetition of the eight-run set was attempted. In Table II, the safety factor, total displacements and steel tension output values are presented along with the experimental schedule. In Table III, all data are transformed to rank values. On a first step, this is done for each quality characteristic individually. On the second step, the ranks are squared and subsequently are added to produce the column SR which represents the final-stage quality metric, i.e. Super-Rank. Since all calculations presented in this work were performed on the statistical software package Minitab

Table II. Orthogonal array L8(27) and quality characteristics data

Exp.	A	B	C	D	E	F	G	Safety factor	Total displacements	Steel tension
1	1	1	1	1	1	1	1	1.213	83.27	1.120
2	1	1	1	2	2	2	2	1.410	12.60	1.334
3	1	2	2	1	1	2	2	0.760	19.09	1.408
4	1	2	2	2	2	1	1	1.534	11.57	0.940
5	2	1	2	1	2	1	2	1.526	10.86	0.379
6	2	1	2	2	1	2	1	0.236	12.65	0.932
7	2	2	1	1	2	2	1	1.720	9.23	0.470
8	2	2	1	2	1	1	2	1.215	36.94	0.699

Table III. Orthogonal array L8(27) and super-ranked data

Exp.	A	B	C	D	E	F	G	RSF	RTDD	RST	SRSF	SRTDD	SRST	SSR	SR
1	1	1	1	1	1	1	1	6	8	6	36	64	36	136	7
2	1	1	1	2	2	2	2	4	4	7	16	16	49	81	4
3	1	2	2	1	1	2	2	7	6	8	49	36	64	149	8
4	1	2	2	2	2	1	1	2	3	5	4	9	25	38	3
5	2	1	2	1	2	1	2	3	2	1	9	4	1	14	2
6	2	1	2	2	1	2	1	8	5	4	64	25	16	105	6
7	2	2	1	1	2	2	1	1	1	2	1	1	4	6	1
8	2	2	1	2	1	1	2	5	7	3	25	49	9	83	5

Notes: RSF = Ranked safety factor data; RTDD = Ranked total displacement difference from nominal value data; RST = Ranked steel tension; SRSF = Squared ranked safety factor data; SRTDD = Squared ranked total displacement difference from nominal value data; SRST = Squared ranked steel tension; SSR = Sums of squared ranks; SR = Super Rank (version 13.32), the Mann-Whitney test is more conveniently applied, in this package, in establishing statistical significance of setting differences for all implicated factors. In Table IV, the super-ranked data is tabulated in terms of median value of factor settings for each factor separately along with the computed significance. The level of significance for this study was set at 0.05. According to this limit, only the factor E ($p = 0.030$) i.e. the angle of excavation seems to influence the concurrent optimization of the three quality characteristics. This result is agreeable to the outcome obtained by employing the common method of the half-normal plot test. Further discussion of nonparametric advantages over the four common graphical tests on unreplicated fractional factorials is taken up in Besseris (2008).

Moreover, the optimum setting occurs at E_2, i.e. for an angle of excavation 4:3. Hence, the response prediction value in this situation will contain the summation of the overall median and the difference between the median value at E_2 and the overall median. This summation results in direct and solely dependence of the prediction value on the median value of E_2. In Tables, V-VII, there are accumulated all factors and their levels for each response separately. Therefore, the optimal value for safety factor is found to be, from Table V, 1.53 with a 95 percent confidence interval of (1.41;1.72), for total displacements the predicted median is 1.22 mm (Table VI) with a 95 percent confidence interval of (0.77 mm; 2.6 mm) and finally for steel tension the corresponding median value is 0.705 T/cm^2 (Table VII) with a 95 percent confidence interval of (0.379 T/cm^2; 1.334 T/cm^2). At this point, it would be interesting to compare these predictions with the optimized one-at-a-time values for each response individually. In Table V, the statistical significance for all seven factors with respect to the safety factor is tabulated for all levels. The optimum setting and the prediction value agrees with the results obtained from considering concurrent optimization above. Also, it is important to note that the safety factor should be maintained at a minimum value of 1.4 as discussed before. According to the estimated confidence interval of safety factor above, this condition is achievable from the present set of data. The same factor setting, E_2, optimizes the total displacement value at 11.21 mm which holds true for the same confidence interval as in the case of simultaneous optimization. This value is well within the maximum allowable value of 50 mm and fairly successful with the internal

Levels	Median ranks	Mann-Whitney test statistic W	p	a
A1	5.5	22	0.312	0.05
A2	3.5			
B1	5.0	19	0.885	0.05
B2	4.0			
C1	4.5	17	0.885	0.05
C2	4.5			
D1	4.5	18	1.000	0.05
D2	4.5			
E1	6.5	26	0.030	0.05
E2	2.5			
F1	4.0	17	0.885	0.05
F2	5.0			
G1	4.5	17	0.885	0.05
G2	4.5			

Table IV. Statistical significance of median differences of super-ranked response per factor

Table V. Statistical significance of median differences for single Safety Factor optimization for each factor

Levels	Median ranks	Median response	95 percent confidence interval	Mann-Whitney test statistic W	*p*	*a*
A1	5.0	1.312	0.760-1.535	19	0.885	0.05
A2	4.0	1.371	0.237-1.721			
B1	5.0	1.312	0.237-1.527	21	0.470	0.05
B2	3.5	1.375	0.760-1.721			
C1	4.5	1.313	1.214-1.721	16	0.665	0.05
C2	5.0	1.144	0.237-1.535			
D1	4.5	1.370	0.760-1.721	17	0.885	0.05
D2	4.5	1.313	0.237-1.535			
E1	6.5	0.987	0.237-1.216	26	0.030	0.05
E2	2.5	1.531	1.411-1.721			
F1	4.0	1.371	1.214-1.535	16	0.665	0.05
F2	5.5	1.085	0.237-1.721			
G1	4.0	1.374	0.237-1.721	17	0.885	0.05
G2	4.5	1.313	0.760-1.527			

Table VI. Statistical significance of median differences for total displacements differences from nominal value for each factor (single response optimization)

Levels	Median ranks	Median response	95 percent confidence interval	Mann-Whitney test statistic W	*p*	*a*
A1	5.0	5.85	1.57-73.27	21	0.470	0.05
A2	3.5	1.76	0.77-26.94			
B1	4.5	2.63	0.86-73.27	19	0.885	0.05
B2	4.5	5.33	0.77-26.94			
C1	5.5	14.77	0.77-73.27	20	0.665	0.05
C2	4.0	2.11	0.86-9.09			
D1	4.0	4.98	0.77-73.27	17	0.885	0.05
D2	4.5	2.63	1.57-26.94			
E1	6.5	18.05	2.65-73.27	26	0.030	0.05
E2	2.5	1.22	0.77-2.6			
F1	5.0	14.26	0.86-73.27	20	0.665	0.05
F2	4.5	2.63	0.77-9.09			
G1	4.0	2.11	0.77-73.27	17	0.885	0.05
G2	5.0	5.85	0.86-26.94			

specification of 10 mm. In terms of the outcomes of the set of trials in Table II this may be received as the most desirable result eventually. The situation is not comparable to the other two responses when it comes to steel tension. According to Table VII, the factor that minimizes statistically steel tension by itself is only factor A, i.e. reinforcement at the factor setting of A_2, i.e. HEB IPB 180. The prediction median of steel tension at this setting is 0.585 T/cm^2 with a 95 percent confidence interval of (0.379 T/cm^2; 0.933 T/cm^2). Thus, there is a 17 percent difference with respect to the optimal value of 0.705 T/cm^2 estimated above. Thus, this is still an excellent overall prediction if one realizes that the quantity we seek to minimize here is within the significance interval for both approaches, concurrently or individually. Specifically, both response predictions include the lower limit of 0.379 T/cm^2. By observing Figure 6, it becomes clear that one of the three responses, the total displacement difference from

Levels	Median ranks	Median response	95 percent confidence interval	Mann-Whitney test statistic W	p	a
A1	6.5	1.228	0.941-1.408	26	0.030	0.05
A2	2.5	0.585	0.379-0.933			
B1	5.0	1.027	0.379-1.334	18	1.000	0.05
B2	4.0	0.82	0.47-1.408			
C1	4.5	0.91	0.47-1.334	18	1.000	0.05
C2	4.5	0.937	0.379-1.408			
D1	4.0	0.796	0.379-1.408	17	0.885	0.05
D2	4.5	0.937	0.700-1.334			
E1	5.0	1.027	0.700-1.408	21	0.470	0.05
E2	3.5	0.705	0.379-1.334			
F1	4.0	0.82	0.379-1.121	15	0.470	0.05
F2	5.5	1.134	0.47-1.408			
G1	3.5	0.937	0.47-1.121	14	0.312	0.05
G2	5.5	1.017	0.379-1.408			

Table VII. Statistical significance of median difference for steel tension for each factor (single response optimization)

the nominal value is not normally distributed. Thus, standard analysis techniques that rely on mean value estimation may not provide an acceptable estimation. The technique outlined in this work relying on non-parametric statistics is suitable for cases where knowledge of the distribution may not be distinguishable for the samples taken. This may turn out to be an added asset of this method. Furthermore, in Appendix 1, a list is provided for Anderson-Darling Normality Test applied to safety factor, total displacements and the steel tension for each factor level separately based on the original values of these responses and not on their respective ranks. It is observed that again the total displacements outcomes pose normality departures for the following factor levels A_1, A_2, B_1, D_1, D_2 and G_1 indicating that handling their group values with normal-related central tendency estimators may not be desirable.

Results gained by any proposed method on the optimization of multiple-characteristics have been accustomed to be compared with the results obtained by the well-accepted desirability-function method (DFM) (Harrington(1965) and, Derringer and Suich(1980)). Actually, the DFM seems to be the only method that has an intimate connection to quality improvement and subsequently to multi-response optimization at large since DFM's originators are in the quality field practitioners themselves. The DFM may even handle the difficult problem of the unreplicated multi-response optimization, though it is not necessarily that will involve the sacred concept of the statistical significance along the way. Owing to the length of this work, we consider only the situation where the recipient of the optimized product has not particular preferences on the individual considered responses. Application of the SR method to the problem where customer preferences enter the optimization mechanism requires a minor modification on the proposed method and it will be treated elsewhere. In this particular case study, trying to compare the results generated by SR as well as with those produced by DFM "head-on" does not do justice of either of the methods. The SR approach is a simple statistical method while the DFM is an empirical deterministic method. Before numerical comparisons are laid out between the two methods, there are some specific features that make each of the two methods distinct. The most important observation is that DFM relies essentially "up-front" on

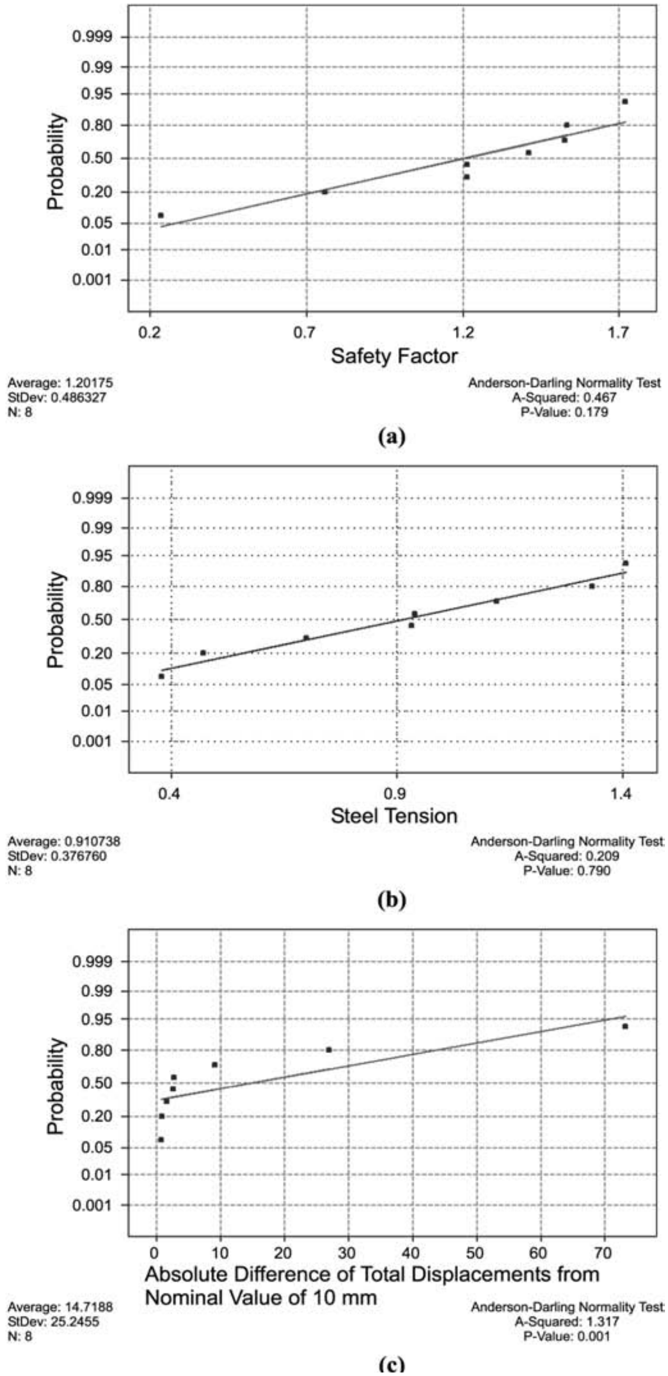

Figure 6.
Normal probability plots for a) safety factor; b) steel tension; and c) absolute difference of total displacement from nominal value of 10 mm

all factors to contribute to a final definite optimized state. In particular, for the case of unreplicated fractional-factorial designs, the response values generated by executing experimental runs will have to surrender their statistical nature and follow a deterministic path to optimization. This way each implicated factor obviously is called upon to provide a share to the optimized solution. Thus, this method would not promote factor elimination in some fashion. On the other hand, for reasonably economical (orthogonal) sampling schemes, the SR method utilizes exact statistics to spot active effects. Without necessarily relying on the premise that only a few effects should be active, SR allocates statistical significance on each factor effect thus setting the stage for subsequent factor elimination. Exact statistics involve full statistical information from all data points thus may be closer to gauge the situation at hand especially for the expected low count sampling. This property stems from the inclusion of the Wilcoxon-Mann-Whitney test in the proposed method thus allowing direct comparison between distributions.

Overall, the SR method may offer a few advantages over the DFM that may prove to be practical and useful to users when they consider the solution of the unreplicated multi-response problem on short experimental schedules as dictated by orthogonal fractional factorial designs. The first advantage is that the proposed method (SR) obviously has low computational complexity. To most cases, the use a calculator will suffice. The second easily discerned advantage of SR over the DFM is that what is offered here is statistical modelling in contrast to deterministic deciphering. To get to the point, getting excellent results with DFM, shop floor managers must be assisted by a statistician and an expert on optimization. This is important because rescaling and weighting individual targets requires very specialized knowledge of the product at hand and the method employed. To this author’s experience eyeing at the turbulent times that modern business is facing today seldom such experts will be rendered. On the other hand, SR may be implemented by anybody with a basic degree that included quantitative methods in business or engineering. Moreover, SR offers relief form the inherent unimodality assumption deeply rooted in DFM which otherwise seldom is checked if followed. SR will work with any distribution. This is particularly welcomed for the economical and small sampling conditions this method is primarily intended to. The SR statistical engine based on Wilcoxon-Mann-Whitney will honour possible distribution peculiarities and limited data and will provide statistical confidence based on exact statistics when each factor setting amasses data points in single-digit count. This last feature is highly desirable and much needed when data point generation really costs or must be produced on very tight schedules. The SR method returns the game of factor selection to the good old statistical elimination strategy not adhering to the DFM credo that a product is as good as its worst performing characteristic. Thus, the SR method may find a niche in decision making under uncertainty while the DFM holds a strong post in decision making under certainty. Furthermore, the SR method allows tradeoffs among various outputs such as those encountered in realistic production environments. This is further facilitated by the introduction of the Euclidean measure for ranked variables to form the composite SR vector. This last attribute also warrants mutual preference independence enforced on the required additive rule. This last property is not always warranted in the DFM. The SR method is also free of possible non-linearities that may be surfaced by the logarithmic transformations in DFM. The point estimate approach evidenced in DFM has been removed permanently from the SR method.

Returning to our case study and loading up the "Response Optimizer" in the DOE menu of MINITAB (Release 15) with the targeted and expected values for the three responses, we relied on intuition to provide a set of logical start-up values for the composite desirability function. In total, three scenarios where examined along with a follow up based on the results gained from the SR method. Since the total displacement optimization issue is fairly well wrapped up about a targeted value and the steel tension can only reach a zero value, which leaves essentially the safety factor for extensive experimentation within the limit this presentation is allowed to evolve. In Figure 7, Scenario A is loaded with the maximum observed safety factor from Table II which is 1.72. The rest of the input values are described numerically in Appendix 2 (Scenario A). The composite desirability is about 0.95 while all factors A-G seems to have their share in contributing to this behaviour. The proposed solution is A_2, B_2, $C_{(1.96),}$ D_1, E_2, F_1, G_1. Since the safety factor is not capped essentially by any value, it was thought to experiment by incrementing its maximum observed value to 1.9 just to examine the composite desirability value to surprisingly tickle down to 0.93 (Figure 8, Scenario B). The new set of factor levels are A_2, B_2, $C_{(1.63)}$, D_1, E_2, F_1, $G_{(1.78)}$ (Appendix 2, Scenario B). Thus, contrasting these two versions lead to variability in C and G as the safety factor was given higher magnitude to evolve. However, according to the SR method this is something that should not occurred statistically since C and G are not active. Next, we are inclined to drop the safety factor maximum value to 1.8 intending to enhance again the composite desirability. This time the value for composite desirability (Figure 8, Scenario C) increases indeed to 0.948 and the optimal factor setting are this time A_2, B_2, $C_{(1.91)}$, D_1, E_2, F_1, $G_{(1.1)}$ (Appendix 2, Scenario C). Again, there is fluctuation originated by factors C and G needing an adjustment to enhance the composite desirability. The surprising result comes when the prediction by the SR method is fed directly to the DFM. The optimal solution according to SR method is only E_2 predicting a safety factor value of 1.53. This causes the composite DFM to crash on 1.0 (Figure 8, Scenario D). Obviously, one may sit down and fiddle around to a perfect scenario with the DFM but the SR method will shed the light in one-step. However, the major discrepancy for the DFM stems from the fact that it was unable to kick out all factors but E (Figure 9). The superiority of SR method in statistical terms over the DFM is well demonstrated on this last observation. Indeed, five of the seven factors (Appendix 2, Scenario D) need to be modulated beyond their set values in order for the DFM to release a "perfect" solution. Overall, the steel tension and the total displacement predicted by the DFM are within the confidence interval set by the SR. DFM required input value of 1.53 for safety factor to predict a value of 1.85 which is not in the set of observed outcomes of Table II. From Table V, it is concluded that the expected dependence of safety factor on C and F anticipated by the DFM it may only be an artefact.

The SR method offers great traceability capabilities. This is easily anticipated due to the structured way of setting and accounting the responses. For example, let us work-out factor E which is the only active. From Table III, the SEN values that comprise level 1 for E are: 83, 105, 136, 139 and the corresponding values for level 2 are: 6, 14, 38, 81. Each of these values carries a triplet of information associated with each of the three responses. For instance, E_1 group SEN value at run 8 is $83 = 25 + 49 + 9$

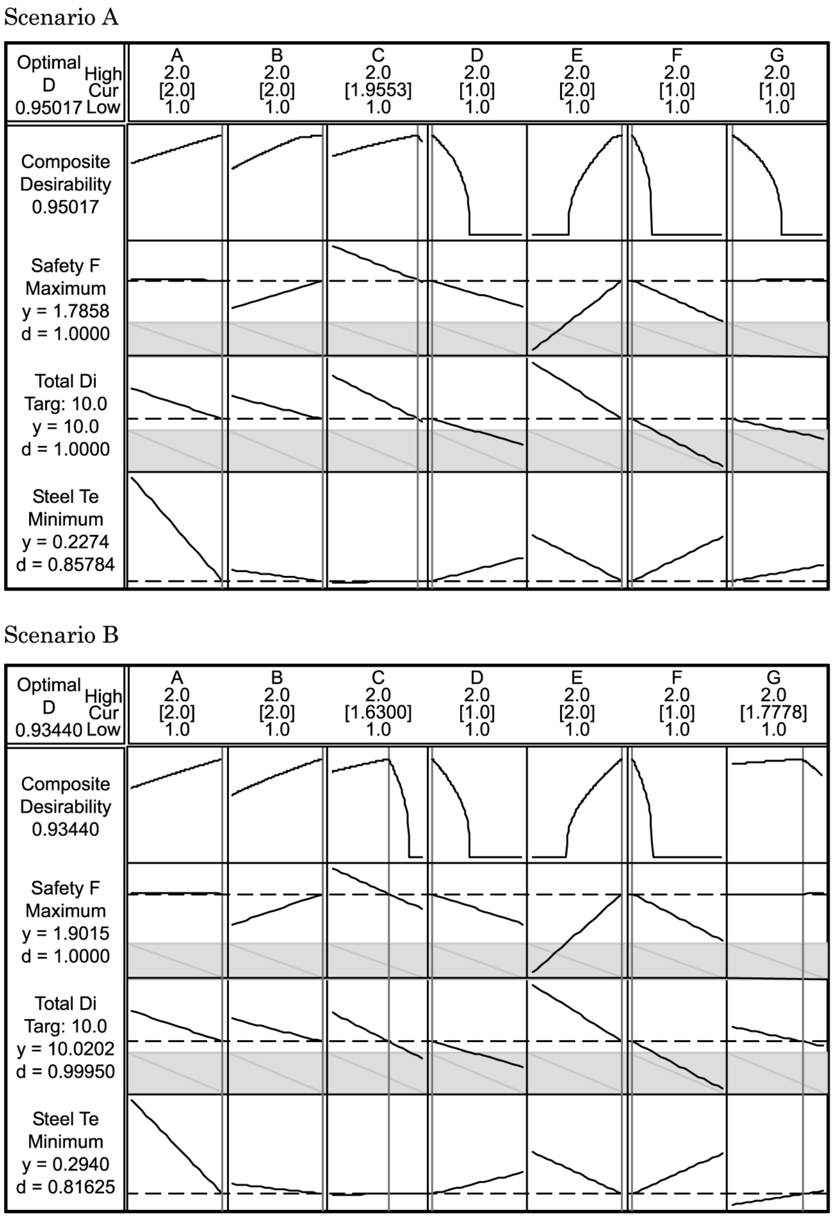

Figure 7.
Graphical solution by the desirability method: scenarios A and B displayed

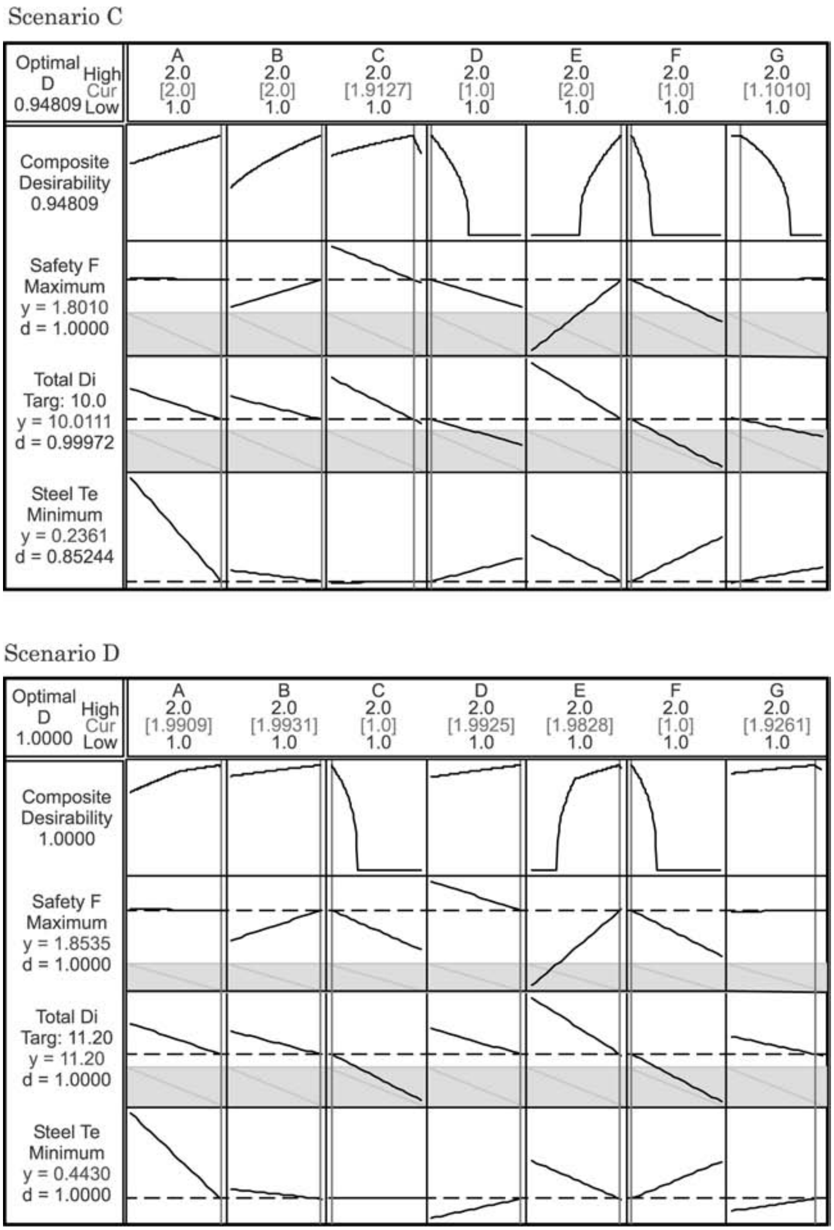

Figure 8.
Graphical solution by the desirability method: scenarios C and D displayed

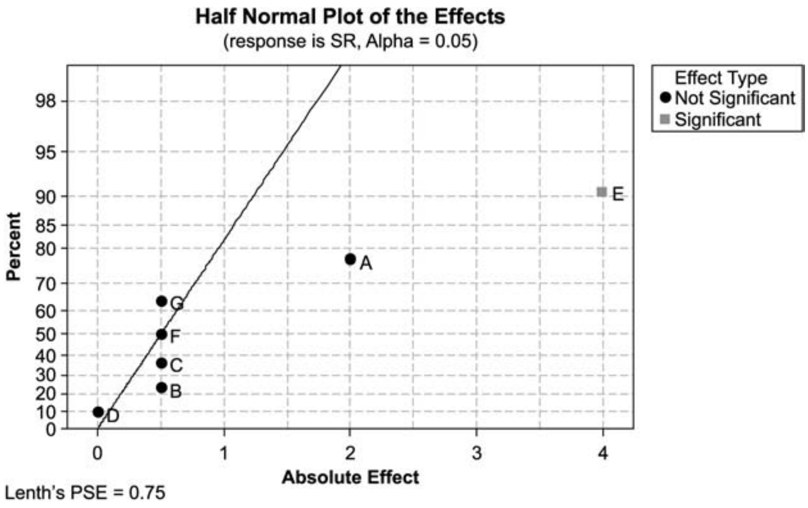

Figure 9. Half-normal plot of the effects based on the SR values

(SRSF + SRTDD + SRST). Therefore, if we want to pinpoint the contributions per response to an active factor setting then the following analysis is conveniently contained (Table III):

Active factor level: group setting SEN values(= SRSF + SRTDD + SRST)

E_1: 83(= 25 + 49 + 9), 105 (= 64 + 25 + 16), 136(= 36 + 64 + 36), 149(= 49 + 36 + 64).

E_2: 6(= 1 + 1 + 4), 14(= 9 + 4 + 1), 38(= 4 + 9 + 25), 81(= 16 + 16 + 49)

which is rearranged as:

(1) E_1:
- SRSF total: 25 + 64 + 36 + 49 = 174 (36.8 percent).
- SRTDD total: 49 + 25 + 64 + 36 = 174 (36.8 percent).
- SRST total: 9 + 16 + 36 + 64 = 125 (26.4 percent).
- Level sum total: 473 (100.0 percent).

(2) E_2:
- SRSF total: 1 + 9 + 4 + 16 = 30 (21.6 percent).
- SRTDD total: 1 + 4 + 9 + 16 = 30 (21.6 percent).
- SRST total: 4 + 1 + 25 + 49 = 79 (56.8 percent).
- Level sum total: 139 (100.0 percent).

From the rank analysis above some practical points become evident. The contribution of SRSF and SRTDD on the active-factor-level sum-total results in equal share for both factor settings. However, at E_1 setting, SRSF and SRTDD lead to higher sum totals which constitutes an undesirable effect since ranks should always be minimized. On the contrary, it is exactly SRSF and SRTDD that maintain the low rank-ability for the E_2 setting. We conclude that SRST is important in minimizing the E_1 outcome in level-sum totals while for E_2 setting its mere consideration constitutes a handicap as it creates the largest part of sum-total augmentation. Thus, no response may be eliminated as weak information shaper comparing to the others.

Discussion

A novel way of handling the general problem of product or process unreplicated multi-response optimization for orthogonal saturated fractional factorial designs has been proposed in this paper. By resorting to a ranking scheme, an attempt has been made to consolidate all implicating quality characteristics to a single-response rank variable. This concept transforms multi-column sets of information to a single column that is easily and conveniently optimized as a new one-response entity. Additionally, this notion that was cultivated herein allows flexibility in combining responses of any nature while at the same time there seems to have the potential to accommodate any number of characteristics. Moreover, this technique was developed to serve as an alternative proposal in processing and analyzing problems in a specific niche of the multi-disciplinary field of industrial experimentation. The need for shortened experimental schedules has been all more often a critical issue nowadays. Budget limitations and expeditious product cycles necessitate drastically reduced trial populations and durations. One proven way to circumvent these restraints is by employing fractional factorial designs during a screening phase of an investigation. Thus, the case study worked out in this report also took in account this need. By relying on standard comparative non-parametric techniques, the method proposed here targets situations where small data samples of any kind of distribution be analyzed effectively. The technique may be worth considering even for unknown types of data distributions since the statistical interpretation is achieved by the all-purpose ranking method of Wilcoxon's or Mann-Whitney's. Allowing intermixing of responses of any origin, which may possess known and unknown data distributions, makes this method unique. A connection between fractional factorial designs, multi-response non-parametric tests and pre-design simulations has been elicited to encompass a statistical viewpoint of many-factor and multi-response relationships with an application to pre-designing quality in construction blueprints.

A case study was selected to demonstrate the proposed method on the complicated and at the same time rarely-published situation where all three optimizing conditions in quality design were present, i.e. "larger-is-better", "smaller-is-better" and "nominal-is-best". The method of individual characteristic ranking leading to a super-rank response assimilates all these three optimizing conditions in a very simplistic and easy to comprehend fashion. This is accomplished by assigning the smaller rank magnitudes to response values that individually are paving their way towards their respective optimizing directions. In this manner, the problem of having to deal with various variables simultaneously, each possibly following its own unit scale, has been convincingly eliminated. Direct normalization of the response variables by their maximum and minimum values is not mandatory when super-ranking is employed in contrast with long-time published methods utilizing the desirability function.

The method outlined here may be extended for cases where there are assigned preferences with respect to the optimized quality characteristics. One such practical proposal may be the introduction of weight factors that reflect the need for preferential treatment of specific responses against each other. Therefore, before forming the super-rank column, each individual response, in rank form, may be multiplied by its respective appropriate weight factor before being encompassed in the terminal form, that of the super-ranked response. However, this is the subject for a forthcoming paper. The use of Analysis of Variance has been a standard and indispensable tool in contrasting factor effects against a response variable. When the size of sampled data is maintained small then assumptions such as normality, randomness and independence of errors and homogeneity of variance may not be easy to be established and alternative methods usually are recommended. When unreplicated and saturated schemes are employed then ANOVA runs out of discriminating power. The only way to handle the multi-response analog problem of the single response unreplicated saturated designs is diverting to the use of DFM and other deterministic techniques such as linear programming, expert systems and so forth. Among these deterministic techniques, DFM is the easiest to implement and thus it has been adopted in many statistical software packages. But methods such as the DFM, is decision making under certainty and alternative methods may need to be searched. Employing non-parametric methods such complications may be overcome and this in turn may be beneficial from a practical perspective. The technique presented in this work has the added advantage that it relies on forcing data homogenization on each engaging response before incurring the multi-response fusion to a single super-ranked response. This may lead us to contemplate that the super-ranking technique may be accessible to more mainstream applications. This is because manufacturing and service environments need daily to solve multi-response optimization problems while it is often certain that they may not always possess the otherwise high-level knowledge of the techniques referred to previously.

As a final note, the case study analyzed here belongs to the fairly new area of pre-designing robust products before even physical experiments have a chance to be performed. DOE in product and process simulations have been a subject only addressed in the past decade and it is expected to find great opportunities for exploration in time to come as economic resources allotted for product development in the competitive institutions dwindle year-after-year. The fact that large-factor pre-design simulations may override actual physical experimentation also has been referred. Thus the proposal made in this article may aid also in this direction too.

References

Ames, A.E., Mattucci, N., MacDonald, S., Szonyi, G. and Hawkins, D.M. (1997), "Quality loss functions for optimization across multiple response surfaces", *Journal of Quality Technology*, Vol. 29, pp. 339-46.

Antony, J. (2000), "Multi-response optimization in industrial experiments using Taguchi's quality loss function and principal component analysis", *Quality and Reliability Engineering International*, Vol. 16, pp. 3-8.

Antony, J. (2001), "Simultaneous optimization of multiple quality characteristics in manufacturing processes using Taguchi's quality loss function", *International Journal of Advanced Manufacturing Technology*, Vol. 17, pp. 134-8.

Antony, J., Anand, R.B., Kumar, M. and Tiwari, M.K. (2005), "Multiple response optimization using Taguchi methodology and neuro-fuzzy based model", *Journal of Manufacturing Technology Management*, Vol. 17, pp. 908-25.

Arvidsson, M. and Gremyr, I. (2008), "Principles of robust design methodology", *Quality and Reliability Engineering International*, Vol. 24, pp. 23-35.

Bates, R.A., Kenett, R.S., Steinberg, D.M. and Wynn, H.P. (2006), "Achieving robust design from computer simulations", *Quality Technology & Quantitative Management*, Vol. 3 No. 2, pp. 161-77.

Berni, R. and Gonnelli, C. (2006), "Planning and optimization of a numerical control machine in a multiple response case", *Quality and Reliability Engineering International*, Vol. 22, pp. 517-26.

Besseris, G.J. (2008), "Analysis of an unreplicated fractional-factorial design using nonparametric tests", *Quality Engineering*, Vol. 20 No. 1, pp. 96-112.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (2005), *Statistics for Experimenters: Design, Innovation, Discovery*, 2nd ed., John Wiley & Sons, New York, NY.

Conover, W.J. (2006), *Practical Nonparametric Statistics*, 3rd ed., Academic Internet Publications, Wellesley, MA.

David, H.A. and Nagarajia, H.N. (2003), *Order Statistics*, 3rd ed., Wiley Series in Probability and Statistics, Wiley-Interscience, New York, NY.

Day, R.W. (1999), *Geotechnical and Foundation Engineering*, McGraw-Hill Professional, London.

Del Castillo, E., Montgomery, D.C. and McCarville, D.R. (1998), "Modified desirability functions for multiple response optimization", *Journal of Quality Technology*, Vol. 28, pp. 337-45.

Derringer, G. and Suich, R. (1980), "Simultaneous optimization of several response variables", *Journal of Quality Technology*, Vol. 12, pp. 214-19.

Dey, A. (1985), *Orthogonal Fractional Factorial Designs*, Halsted Press, New York, NY.

Elsayed, E.A. and Chen, A. (1993), "Optimal levels for process parameters for products with multiple characteristics", *International Journal of Production Research*, Vol. 31 No. 5, pp. 1117-32.

Harrington, E.C. (1965), "The desirability function", *Industrial Quality Control*, Vol. 21, pp. 494-8.

Hazod, W. (2001), *Stable Probability Measures on Euclidean Spaces and on Locally Compact Groups*, 1st ed., Springer, Dordrecht.

Hsieh, K.-L. and Tong, L.-I. (2001), "Optimization of multiple quality responses involving qualitative and quantitative characteristics in IC manufacturing using neural networks", *Computers in Industry*, Vol. 46 No. 1, pp. 1-12.

Hsieh, K.-L., Tong, L.-I., Chiu, H-P. and Yeh, H-Y. (2005), "Optimization of a multi-response problem in Taguchi's dynamic system", *Computers & Industrial Engineering*, Vol. 556, p. 571.

Huber, P.J. (2003), *Robust Statistics*, 1st ed., Wiley Series in Probability and Statistics, Wiley-Interscience, New York, NY.

Jayaram, J.S.R. and Ibrahim, Y. (1999), "Multiple response robust design and yield maximization", *International Journal of Quality & Reliability Management*, Vol. 16 No. 9, pp. 826-37.

Kackar, R.N. (1985), "Off-line quality control, parameter design and the Taguchi method", *Journal of Quality Technology*, Vol. 17 No. 4, pp. 176-209.

Kenett, R.S. and Steinberg, D.N. (2006), "New frontiers in the design of experiments", *Quality Progress*, August, pp. 61-5.

Khuri, A.I. and Conlon, M. (1981), "Simultaneous optimization of multiple responses represented by polynomial regression functions", *Technometrics*, Vol. 23 No. 4, pp. 363-75.

Kim, K-J. and Lin, D.K.J. (2000), "Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions", *Applied Statistics*, Vol. 49 No. 3, pp. 311-25.

Kros, J.F. and Mastrangelo, C.M. (2001), "Comparing methods for the multi-response design problem", *Quality and Reliability Engineering International*, Vol. 17, pp. 323-31.

Kumar, P., Barua, P.B. and Gaindhar, J.L. (2000), "Quality optimization (multi-characteristics) through Taguchi's technique and utility concept", *Quality and Reliability Engineering International*, Vol. 16, pp. 475-85.

Liao, H.-C. (2003), "Using PCR-TOPSIS to optimize Taguchi's multi-response problem", *International Journal of Advanced Manufacturing Technology*, Vol. 22, pp. 649-55.

Liao, H.-C. and Chen, Y.-K. (2002), "Optimizing multi-response problem in the Taguchi method by DEA based ranking method", *International Journal of Quality & Reliability Management*, Vol. 19 No. 7, pp. 825-37.

Lu, D. and Antony, J. (2002), "Optimization of multiple responses using a fuzzy-rule based inference system", *International Journal of Production Research*, Vol. 40, pp. 1613-25.

Mann, H.B. and Whitney, D.R. (1947), "On a test of whether one of two random variables is stochastically larger than the other", *The Annals of Mathematical Statistics*, Vol. 18 No. 1, pp. 50-60.

Montgomery, D.C. (2004), *Design and Analysis of Experiments*, 5th ed., Wiley, New York, NY.

Moorthy, S. (1998), *Visionary Manufacturing Challenges for 2020*, National Academy Press, Washington, DC.

Phadke, M.S. (1989), *Quality Engineering Using Robust Design*, Prentice-Hall, Englewood Cliffs, NJ.

Purcell, J. (1999), "Simulation software cuts time and expense", *Control Magazine*, October, pp. 11-19.

Reddy, P.B.S., Nishina, K. and Babu, A.S. (1997), "Unification of robust design and goal programming for multi-response optimization – a case study", *Quality and Reliability Engineering International*, Vol. 13, pp. 371-83.

Reddy, P.B.S., Nishina, K. and Babu, A.S. (1998), "Taguchi's methodology for multi-response optimization: case study in the Indian plastics industry", *International Journal of Quality & Reliability Management*, Vol. 15 No. 6, pp. 646-68.

Robinson, T.J., Borror, C.M. and Myers, R.H. (2004), "Robust parameter design: a review", *Quality and Reliability Engineering International*, Vol. 20, pp. 81-101.

Romano, D., Varetto, M. and Vicario, G. (2004), "Multiresponse robust design: a general framework based on combined array", *Journal of Quality Technology*, Vol. 36, pp. 27-37.

Ross, P.J. (1988), *Taguchi Techniques for Quality Engineering*, McGraw-Hill, New York, NY.

Sacks, J. and Welch, W.J. (1991), "A system for quality improvement via computer experiments", *Communications in Statistics*, Vol. 20, pp. 477-95.

Sacks, J., Schiller, S.B. and Welch, W.J. (1989b), "Design for computer experiments", *Technometrics*, Vol. 31, pp. 41-7.

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989a), "Design and analysis of computer experiments", *Statistical Science*, Vol. 4 No. 4, pp. 409-23.

Schrage, M. (2000), *Serious Play: How the World's Best Companies Simulate to Innovate*, Harvard Business School Press, Boston, MA.

Taguchi, G., Chowdhury, S. and Taguchi, S. (2000), *Robust Engineering – Learn How to Boost Quality while Reducing Costs and Time to Market*, McGraw-Hill, New York, NY.

Tang, L.C. and Xu, K. (2002), "A unified approach for the dual response surface optimization", *Journal of Quality Technology*, Vol. 34 No. 4, pp. 437-47.

Taper, M.L. and Lele, S.R. (2004), *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations*, University of Chicago Press, Chicago, IL.

Telis, E.S.N. and Besseris, G.J. (2007), "Taguchi optimization of displacements during the design phase of a construction project", paper #27, paper presented at the 2nd International Conference on Construction Engineering and Management, Technical Session 5D, Meritus Mandarin Singapore, Singapore, March 1-2.

Tong, L.-I. and Hsieh, K.L. (2000), "A novel means of applying neural networks to optimize the multiple response problem", *Quality Engineering*, Vol. 13 No. 1, pp. 11-18.

Tong, L.-I. and Wang, C.H. (2002), "Multi-response optimization using principal component analysis and grey relational analysis", *International Journal of Industrial Engineering*, Vol. 9 No. 4, pp. 343-50.

Tong, L.-I., Su, C.-T. and Wang, C.-H. (1997), "The optimization of multi-response problems in the Taguchi method", *International Journal of Quality & Reliability Management*, Vol. 14 No. 4, pp. 367-80.

Vining, G. (1998), "A compromise approach to multiresponse optimization", *Journal of Quality Technology*, Vol. 30 No. 3, pp. 309-13.

Wilcoxon, F. (1945), "Individual comparisons by ranking methods", *Biometrics Bulletin*, Vol. 1 No. 6, pp. 80-3.

Wood, D. (2004), *Geotechnical Modeling*, Spon Press, London.

Wu, C.F.J. and Hamada, M. (2000), *Experiments: Planning, Analysis and Parameter Design Optimization*, John Wiley & Sons, New York, NY.

Wurl, R.C. and Albin, S.L. (1999), "A comparison of multiresponse optimization: sensitivity to parameter selection", *Quality Engineering*, Vol. 11 No. 3, pp. 405-15.

Further reading

Moorthy, S. (1999), *Defense Manufacturing in 2010 and Beyond: Meeting the Changing Needs of National Defense*, National Academy Press, Washington, DC.

Appendix 1

Anderson-Darling Normality Test for Safety Factor, Total Displacements and Steel Tension for each factor level considered in Table AI.

Table AI.

Levels	Safety factor A^2	(p)	Total displacements A^2	(p)	Steel tension A^2	(p)
A1	0.272	(0.445)	0.661	(0.025)	0.228	(0.575)
A2	0.342	(0.270)	0.628	(0.032)	0.220	(0.604)
B1	0.458	(0.114)	0.786	(0.010)	0.250	(0.512)
B2	0.209	(0.652)	0.349	(0.256)	0.198	(0.702)
C1	0.380	(0.204)	0.359	(0.238)	0.213	(0.635)
C2	0.348	(0.258)	0.491	(0.089)	0.281	(0.417)
D1	0.205	(0.672)	0.619	(0.034)	0.317	(0.324)
D2	0.453	(0.118)	0.757	(0.012)	0.317	(0.323)
E1	0.353	(0.248)	0.355	(0.245)	0.158	(0.861)
E2	0.296	(0.376)	0.183	(0.769)	0.270	(0.451)
F1	0.552	(0.056)	0.392	(0.186)	0.181	(0.775)
F2	0.202	(0.682)	0.347	(0.261)	0.284	(0.409)
G1	0.342	(0.269)	0.751	(0.013)	0.400	(0.174)
G2	0.282	(0.416)	0.391	(0.187)	0.317	(0.324)

Appendix 2

Response optimization scenarios by the desirability function method (MINITAB, Release 15 output) (Figures A1-A4).

```
Parameters

                      Goal     Lower  Target  Upper   Weight  Import
Safety Factor         Maximum    1.4    1.72   1.72       1       1
Total Displacement    Target     5.0   10.00  50.00       1       1
Steel Tension         Minimum    0.0    0.00   1.60       1       1

Global Solution

A   =         2
B   =         2
C   =         1.95
D   =         1
E   =         2
F   =         1
G   =         1

Predicted Responses
Safety Factor         =    1.785    desirability =   1.000
Total Displacement    =   10.000    desirability =   1.000
Steel Tension         =    0.227    desirability =   0.857
Composite Desirability = 0.950
```

Figure A1. Response optimization scenario A

Parameters

	Goal	Lower	Target	Upper	Weight	Import
Safety Factor	Maximum	1.5	1.9	1.9	1	1
Total Displacement	Target	5.0	10.0	50.0	1	1
Steel Tension	Minimum	0.0	0.0	1.6	1	1

Global Solution

```
A   =     2
B   =     2
C   =     1.63
D   =     1
E   =     2
F   =     1
G   =     1.78
```

Predicted Responses

```
Safety Factor        =   1.901    desirability =  1.000
Total Displacement   =  10.020    desirability =  0.999
Steel Tension        =   0.294    desirability =  0.816
```

Composite Desirability = 0.934

Figure A2.
Response optimization scenario B

Parameters

	Goal	Lower	Target	Upper	Weight	Import
Safety Factor	Maximum	1.5	1.8	1.8	1	1
Total Displacement	Target	5.0	10.0	50.0	1	1
Steel Tension	Minimum	0.0	0.0	1.6	1	1

Global Solution

```
A   =     2
B   =     2
C   =     1.91
D   =     1
E   =     2
F   =     1
G   =     1.10
```

Predicted Responses

```
Safety Factor        =   1.80    desirability =  1.000
Total Displacement   =  10.01    desirability =  0.999
Steel Tension        =   0.23    desirability =  0.852
```

Composite Desirability = 0.948

Figure A3.
Response optimization scenario C

Parameters

	Goal	Lower	Target	Upper	Weight	Import
Safety Factor	Maximum	1.40	1.53	1.53	1	1
Total Displacement	Target	5.00	11.20	50.00	1	1
Steel Tension	Minimum	0.00	0.71	1.60	1	1

Global Solution

A	=	1.99
B	=	1.99
C	=	1
D	=	1.99
E	=	1.98
F	=	1
G	=	1.92

Predicted Responses

Safety Factor	=	1.85	desirability =	1.000
Total Displacement	=	11.20	desirability =	1.000
Steel Tension	=	0.44	desirability =	1.000

Composite Desirability = 1.000

Figure A4.
Response optimization scenario D

Corresponding author
George J. Besseris can be contacted at: besseris@teipir.gr

IJQRM
26,6

614

Received 29 January 2008
Revised 5 December 2008
Accepted 23 February 2009

Effect of individual components on system's reliability

A case of web-based US Federal Highway Administration project recommendation and approval software

Mark R. Chandler
US DOT-FHWA Wisconsin Division, Madison, Wisconsin, USA, and
M. Affan Badar
Department of ECMET, Indiana State University College of Technology, Terre Haute, Indiana, USA

Abstract

Purpose – The purpose of this study is to examine the effect of individual components' reliability on a system's reliability. The system refers to the Financial Management Information System (FMIS), the US Federal Highway Administration's (FHWA) web-based project approval and tracking software. Its components are 61 project information fields.

Design/methodology/approach – The analysis would view each highway project-funding request as an activity with cycle-dependent performance for which success probability can be calculated as Reliability, R. The reliability analysis of the 61 FMIS fields results in a series system with R_{sys} the "estimated reliability" of finding "true" values in all 61 information fields during one highway-related project funding authorization review.

Findings – Of an estimated 200 projects approved, there was previously estimated a 50 percent to 80 percent unreliability rate, while the study found an unreliability rate of approximately 80 percent.

Research limitations/implications – Owing to the nature of federal government software, data can be very difficult to acquire in this working environment, but a simple calculation was relatively successful in confirming the "estimated reliability" of finding "true" values and showing how the reliability could dramatically decrease.

Originality/value – The paper contributes to the applicability of reliability analysis to project approval software, showing the progression from estimated data to bounding the estimate using reliability theory.

Keywords Worldwide web, Management information systems, Computer software, Project finance, Motorways, United States of America

Paper type Case study

Introduction

This paper focuses on the reliability of project information fields in the Financial Management Information System (FMIS), the US Department of Transportation (DOT) Federal Highway Administration's (FHWA) web-based project approval and tracking software, which in FY 2007 was used to approve approximately $40.9B US highway

The authors acknowledge the support provided by the Wisconsin Division Office of the US Department of Transportation Federal Highway Administration.

International Journal of Quality & Reliability Management
Vol. 26 No. 6, 2009
pp. 614-628

0265-671X
DOI 10.1108/02656710910966156

budget and approximate \$716.8M Wisconsin state share. This research was selected after a review of Kales (1998) nomenclature of reliability parameters. This analysis would view each highway project funding request evaluation as a performance activity with cycle-dependent performance as a one-shot example for which success probability can be calculated as Reliability, R. Reliability is directly linked to the FMIS qualities of accuracy and completeness (Parssian *et al.*, 2004). This paper contributes to the applicability of reliability analysis to specific project approval software (Noy, 2003), FMIS, showing the progression from estimated, to bounding the estimate using reliability theory, to future efforts to replicate the analysis with field data.

Issues related to software and user interface have been addressed by several authors. This includes data and information quality, and the data quality characteristics of accuracy and completeness (Sen, 2001; Parssian *et al.*, 2004), and the broader examination of information systems development involving web services, networks, databases, and customized software applications (Ravichandran and Rai, 1999/2000; Noy, 2003; Jain and Priya, 2005). This is also considered by Madu (2005), in an examination of the reliability literature, referencing reliability, availability and maintainability (RAM).

Background

FMIS – a web-based project software environment – is accessed by the over 50 States (and other) Department of Transportation (DOT) in the USA, the US Department of Transportation (DOT) and the Federal Highway Administration (FHWA), and other stakeholders. One FHWA document (Federal Highway Administration, 2000) states the FMIS "database was designed to keep careful track of funds approved and expended on all Congressional appropriations related to the Highway Trust Fund", and facilitate FHWA project oversight (US Government Accountability Office, 2005). This analysis is restricted to the Wisconsin Department of Transportation (WisDOT) and the Wisconsin Division of the FHWA. Project recommendation and approval via this web-based tool is a critical success activity in the FHWA.

The projects tracked in FMIS are done so with the intention by the State Department of Transportation (WisDOT) to be approved by the federal oversight agency, the Federal Highway Administration (FHWA). The projects may be approved if all relevant procedures have been followed and the FMIS information fields are correct and/or accurate. Missing or incorrect information can have a drastic impact on the project approval request, resulting in no approval for the project. This lack of approval – which is equivalent to a project refusal – is not desired by the state department of transportation or the FHWA, and the FMIS project tracking and approval software is intended to facilitate project approval. However, this analysis shows that significant reliability issues exist with project information fields, leading to routine project refusals, which are usually temporary and very infrequently permanent. This lends itself to a concept referred to later in the paper, that of project information repairability.

For local users there is experience and evidence that the software and interface are typically difficult and cumbersome. The FMIS user would also typically review a project's information and find there are significant information deficiencies that prevent the recommending engineer from completing the initial intention of recommending the project for approval during the project funding authorization review.

Consequently, the engineer/user will have to either leave the computer screen open and look for the needed information or perhaps call and/or e-mail someone. Further, the engineer/user will have to write a note in the Division Office remarks field, indicating that work had been done on the project recommendation, but had to ask for further information and/or changes and could not sign to recommend at this time, i.e. the highway-related project funding authorization review attempt had failed.

Use of the FMIS software has lead to a perception that the likelihood of success with any highway-related project funding authorization review attempt is relatively unlikely. For this paper, the reliability analysis of this process shows the series elements of this process – the FMIS project approval information fields – have failed (Kales, 1998). To document this analysis a series of screen captures have been made, where the first author makes an estimate of the number of fields that have critical information, and the chance the field would contain correct and/or sufficient information, based on work experience with the software. This is a determination of accuracy and completeness (Park and Sanders, 2007; Parssian *et al.*, 2004). This is followed-up with data collection on the use of the software by the authors, to confirm the analysis from the estimated successes. If the field has missing or incorrect information, the reliability consequence is the series element (i.e. information field) fails. This situation somewhat blurs the statement by Sen (2001) that while data quality can be made certain, information quality is often hard to control.

The specific process of recommending authorizations for projects requires two password protected sign-in procedures. The first level is logging onto the Federal Highway Administration computer, and the second is signing into the FMIS electronic environment, shared by FHWA, WisDOT and others. Both systems are robust, but there are sometimes issues with each. There can be power or FHWA system failures or password problems that keep the engineer user from logging onto a computer. There can also be FMIS issues, such as too many users or separate password issues that keep a user from signing into FMIS. The standard procedure is to print out a WisDOT regional report for the projects being reviewed.

For this analysis the 3rd project in Figure 1 is selected, Project ID #2007(010), which is considered by the first author to be fairly typical of those projects shown. Initially, this first screen could have problems with the six of the eight fields: Project ID, State

Figure 1.
Wisconsin Department of Transportation (WisDOT) Southwest region projects

Project #, Signature Needed, Project Oversight, Federal Funds Change + (or -) and Pending Actions. Problems at this level are rare, and would be followed up at the project information screen level, with the complete Financial Management Information System (FMIS) – the Federal Highway Administration's (FHWA) web-based project recommendation and approval software – shown in three successive screen captures, in Figures 2-4. Note in Figures 2-4 the text box and number (i.e. [#]), indicates the series element number found in Table I (1 to 61).

To further frame the use of FMIS a field operations engineer is not concerned with all fields of information. WisDOT staff could be interested in some fields, and the financial specialists who approve the projects after a field operations engineer or user recommends a project for approval – by electronically signing the web-based document, a signature process with a 3rd and 4th password – has some other fields of interest. The areas and related fields are listed in Table I. For this FMIS reliability

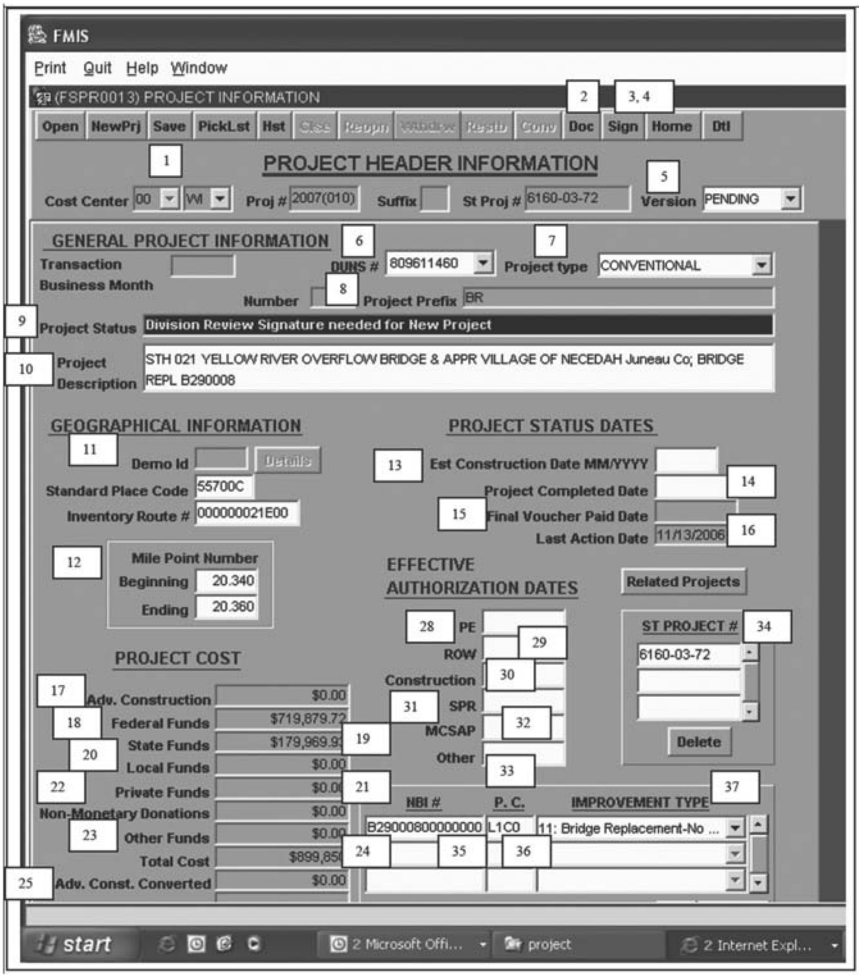

Figure 2.
First screen capture of project 2007(010) from FMIS, the FHWA's web-based project recommendation and approval software

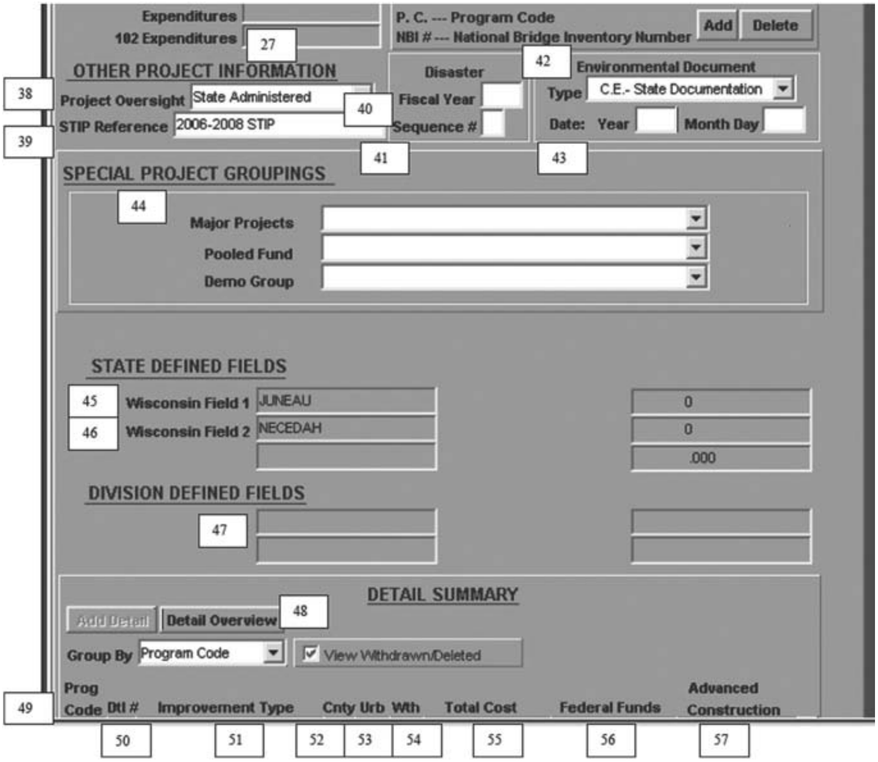

Figure 3.
Second screen capture of project 2007(010) from FMIS

analysis typical experience in the last two years has resulted in an estimate of approved or recommended for approval 200 projects, which was the baseline estimate used in this case study. The time and resources needed to confirm this estimate of the number of FMIS projects approved by the first author from FY2005 to FY2008 would be a significant burden on office resources, and unlikely to be supported.

Methodology

In this analysis the reliability of the elements of a number of FMIS information fields that must be reviewed by a FHWA Wisconsin Division field operations engineer or other user is examined. These fields, as shown in Figures 2-4, can be approximated as being either true or false. There can be a number of reasons why a field is considered false:

(1) no information in the field;

(2) incorrect information in the field;

(3) the field does not have current information;

(4) there is not sufficient information in the field;

(5) the field has a stop condition specified that has to be cleared and is considered a false field until the resolution; and

(6) other.

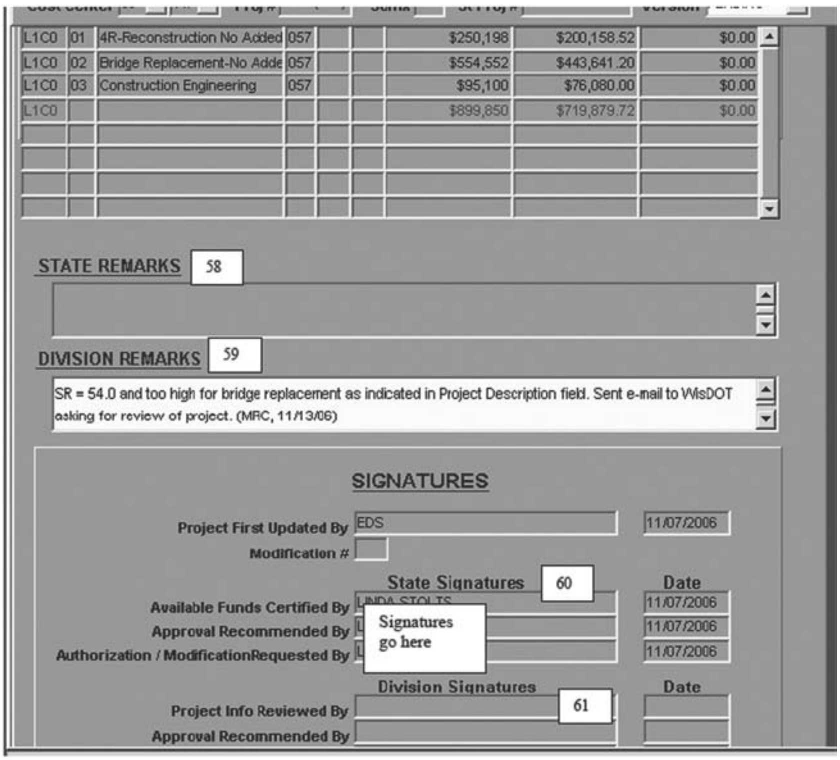

Figure 4.
Third screen capture of project 2007(010) from FMIS

The reliability analysis of the 61 FMIS fields results in a series system (Kales, 1998), with each Field name given a corresponding number of item reliability estimate (Kales, 1998). This follows Kales "best estimate of the Reliability" $R = S/N$, where S is the estimated or sampled number of failures and N is the number of approval/recommendations initiated in FMIS:

$$R = P(S) = S/N \tag{1}$$

where:

$P(S)$ = probability of success;

S = number of successes;

N = number of trials (either success or failure).

As previously stated, following Kales (1998) nomenclature of reliability parameters, this analysis will view each highway-related project funding authorization review as an item performance with cycle-dependent performance, and a one-shot example for which success probability can be calculated as Reliability, R. This reliability modeling for the FMIS web-based software can be viewed as a series system, composed of 61 elements – in this case separate data fields – the failure (i.e. lack of information or

Table I.
R_i and R_{sys} calculation for current information fields

Element #	Area name	Field name	Number of items	S_i	R_i	Comments
1	Project information	Save	1	199	0.9950	Almost always works
2		Doc	1	198	0.9900	Almost always works, but can be a problem to get hardcopy if fails
3		Sign	3rd password	198	0.9900	Almost always works, but password can be a problem
4			4th password	198	0.9900	Almost always works, but password can be a problem
5	Project header information	Version	1, 2 and occasionally more	200	1.0000	
6	General project information	Duns #	1	199	0.9950	Should always be right
7		Project type	5: conventional, emergency relief, demo, STP, SIB	180	0.9000	
8		Project prefix	Approximately 10: STP, NH, IM, BR, BH, MG, OTA, demo, earmarks, CMAQ	180	0.9000	
9		Project Status	2: division review or authorization	199	0.9950	
10		Project description	Field will only hold approximately 100 characters; need location, local government unit, county, type of activity	150	0.7500	This field almost always has insufficient information, but extra required information can be put in the Division Remarks field to meet minimum project description requirements by reviewing field operations engineer
11	Geographical information	Demo Id	1, the assigned #	200	1.0000	Not critical
12		Mile point number	2, the beginning and ending mile points	200	1.0000	Not critical, but lack might be questioned
13	Project status dates	End construction date	1, the appropriate date	199	0.9950	
14		Project completion date	1, the appropriate date	199	0.9950	
15		Final voucher paid date	1, the appropriate date	199	0.9950	

(*continued*)

Element #	Area name	Field name	Number of items	S_i	R_i	Comments
16		Last action date	1, the appropriate date	200	1.0000	Very important field to see when last action took place among several participants in WisDOT and FHWA
17	Project cost	Adv. construction	1, the appropriate amount	200	1.0000	
18		Federal funds	1, the appropriate amount	180	0.9000	Significant information and typically a percent, such as 80 percent, 90 percent or 100 percent
19		State funds	1, the appropriate amount	180	0.9000	Significant information and typically a percent, such as 20 percent, 10 percent or 0 percent
20		Local funds		199	0.9950	Could be important
21		Private funds		200	1.0000	Could be important
22		Non-monetary donations		200	1.0000	
23		Other funds		200	1.0000	
24		Total cost		195	0.9750	Could be important
25		Adv. const. converted		199	0.9950	
26		Expenditures		200	1.0000	
27		102 Expenditures		200	1.0000	
28	Effective authorization dates	PE		195	0.9750	Important
29		ROW		199	0.9950	Important
30		Construction		195	0.9750	Important
31		SPR		200	1.0000	
32		MCSAP		200	1.0000	
33		Other		200	1.0000	
34	Related projects	ST project #		200	1.0000	Surprisingly, this is not important
35	NBI #		1 #	180	0.9000	Important
36	P.C.			180	0.9000	
37	Improvement type		1 or 2	180	0.9000	Important
38	Other project information	Project oversight	4: Full oversight (NHS), Full oversight (non-NHS), State administered, other	200	1.0000	
39		STIP reference		150	0.9500	Important

(*continued*)

Table I.

Table I.

Element #	Area name	Field name	Number of items	S_i	R_i	Comments
40	Disaster	Fiscal year		200	1.0000	Could be important; but is very rare
41		Sequence #		200	1.0000	
42	Environmental document	Type	5: EIS, C.E. – Programmatic, C.E. – State Documentation, C.E. – FHWA Documentation, EA, FONSI	190	0.9500	Important; this information frequently add this "required' information, making an online repair
43		Date	Year & Month	200	1.0000	
44	Special project groupings			200	1.0000	Not important
45	State defined fields	Wisconsin field 1	72: County	190	0.9500	Very important
46		Wisconsin field 2	1780: Local unit of government	180	0.9000	Very important
47	Division defined fields			200	1.0000	Surprisingly, this is not important; seems to be a placeholder in the software
48	Detailed summary	Detail overview (a button to open 2nd window with approximately 33 fields)		199	0.9950	The separate fields could be important; but this is very rare; this is summary information from main screen and special information for others besides a field operations engineer
49		Program code		190	0.9500	Depending on the Program code for each improvement type, there could be a matrix of data, with each cell being important. Each program code row can be clicked on to open another window with approximately 39 fields. As stated above, the separate fields could be important; but this is very rare; this is summary information from main screen and special information for others besides a field operations engineer
50		DTL #		200	1.0000	

(*continued*)

Element #	Area name	Field name	Number of items	S_i	R_i	Comments
51		Improvement type		196	0.9800	There are approximately ten improvement types, and these are important. Some have specific $ percent that have to be checked or researched
52		Cnty		200	1.0000	
53		Urb		200	1.0000	
54		Wth		200	1.0000	
55		Total cost		196	0.9800	
56		Federal funds		190	0.9500	While similar to earlier field, there are differences
57		Advanced construction		190	0.9500	While similar to earlier field, there are differences
58	State remarks			190	0.9500	Could be important; but is very rare
59	Division remarks			180	0.9000	Important
60	Signatures	State signatures	3, automatically adds date	200	1.0000	Very important; currently one person's electronic signature fills all three fields. This is being changed to provide fiscal controls
61		Division signatures	2, automatically adds date	198	0.9900	Very important; one field operations engineer's electronic signature fills two fields
				$R_{sys} = \prod_{i=1}^{n} R_i$	0.1788	Or approximately an 18 percent expected reliability
				$\bar{X} = \frac{\sum_{i=1}^{n} R_i}{N}$	193.7541 $\approx$ 194	The mean of the 61 cells in column R_i

Note: $n = 61$, $N = 200$

Table I.

wrong information, or some other condition, such as a mistake or a critical time had not been reached) of any of which will cause a system failure. In this case the failure would be that the intended project funding authorization review of a highway project in FMIS could not be completed, further information and/or changes would have to be made (i.e. repairability), and the highway project would be delayed. Again, after Kales (1998):

$$R_{sys} = R_1 \times R_2 \times \ldots \times R_n = \prod_{i=1}^{n} R_i \, for i = 1, \ldots, n \tag{2}$$

where:

$R_i \quad = S_i/N$

and for this project, $n = 61$, $N = 200$.

The authors suggest that based on the project tracking software and FMIS parameters, one may reasonably assume the 61 components – or FMIS information fields – are statistically independent for the purposes of this analysis (Kales, 1998, pp. 297, 312-13). To make the logic and calculation more aligned with the specifics of this work related activity, the separate 61 elements – in this case separate data fields, S_i – were estimated in Table I. This approach will allow the rapid comparison of reliabilities of the fields.

Results

The 61 fields (i.e. element #, column 1 of Table I) for this analysis and related number of successes (S_i) is estimated by the first author based on work experience, with Figures 2-4 translated into Table I, and along with $n = 61$ and $N = 200$, used to calculate their respective R_i. From these individual R_i, the system's reliability R_{sys} can be obtained from Equation (2). R_{sys} is the "estimated reliability" of finding "true" values in all 61 information fields during one project funding authorization attempt in FMIS.

Discussion

For this analysis the first author estimates of 200 projects considered for approval and subsequently signed to recommend for approval, there was an estimated 50 percent to 80 percent unreliability rate for approval in the FMIS software on the first attempt. In particular, the level of government and county in either the Project Description field or the Division Remarks field would be entered incorrectly or missing by the FMIS user. Of the other projects that required further effort, there would be approximately 75 percent or more (i.e. unreliability of 25 percent or less) that could be approved on the 2nd attempt, after the engineer/user and/or others added information to the required fields or changed the information in the required fields. The remaining projects would take a 3rd or more attempts and would require more time and effort, including examination of other documents, a few to many e-mail exchanges, and one to several phone calls.

As part of this analysis, a spreadsheet was used to manage the calculations needed to provide estimated reliabilities, from Equation (2). The 61 separate R_i values in Table I, along with the total system $R_{sys} = 0.1788$, showing that about 18 percent of the time the attempted authorization/recommended for approval process would be successful. This was "in the ball park" of the first author's – as a local user –

estimates. As a further analysis, the S_i values were reduced incrementally by 1, 10 and 20, which would be what would happen if more of the fields (i.e. independent elements, in this analysis) were not correct. This would have the effect of decreasing the reliability, and increasing the unreliability (Kales, 1998):

$$R + Q = 1 \tag{3}$$

where:

R = Reliability; and

Q = Unreliability

The results of this spreadsheet calculation are shown in Table II, showing the reliability, R, from the initial success estimate (i.e. each field was acceptable and/or correct and needed no further modification) S_i values and the S_i values incrementally decreased. This would indicate more problems with the information fields, and equivalently, the series elements would have failed more frequently. Note that the decrease in success corresponds to 1 less, to 10 less, to 20 less results in a significant decrease in system reliability, corresponding to an increase in unreliability, which is the probability of the lack of success. This simple calculation was relatively successful in showing how the reliability could dramatically decrease. This is a process showing evidence that the FMIS reliability is relatively "good" or projects would very rarely be approved on the first try in FMIS.

As a check of this logic shown in Table II note that a rough approximation of the review of FMIS information fields is an equivalent series, shown in Figure 5.

And if $R_i = 0.97$ (i.e. 194 times out of 200 estimated, the field is found correct) for all 61 elements:

$$R_{sys} = R_1 \times R_2 \times \ldots \times R_n = \prod_{i=1}^{n} R_i = 0.9761 = 0.156 = 0.9761 = 0.156$$

approximately 16 percent, which is close to the 17.88 percent result in the calculation summarized in Table II.

It is important to note only a small number initial success estimates were below 190 out of 200, with S_i maximum value = 200 and minimum value = 150 in Table I. This indicates a relatively high level of reliability of most fields, which are the equivalent

	R_{sys}	Approx. R_{sys}, as percent
Initial S_i	0.1788	17.88
$S_i - 1$	0.1305	13.05
$S_i - 10$	0.0071	0.71
$S_i - 20$	0.0002	0.02

Table II. Reliability from the initial S_i values and the S_i values incrementally decreased

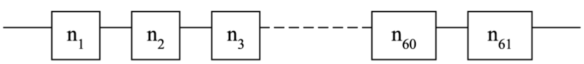

Figure 5. Equivalent series

series elements in this analysis. There were a significant number of S_i information fields that were estimated at 200 because these fields were found to be always correct. There were only a relatively few fields that seemed to have significant problems that would lead to initially lower successes (i.e. $S_i \ll N = 200$).

Conclusion and recommendation

This paper contributes to the applicability of reliability analysis to project approval software, showing the progression from estimated, to bounding the estimate using reliability theory, to future efforts to replicate the analysis with field data. The reliability analysis resulted in the series system calculation of R_{sys}, the "estimated reliability" of finding "true" values in all 61 information fields during one project funding authorization attempt in FMIS. The resulting reliability estimate was 0.1788, or almost 18 percent, of a field operations engineer or software user being able to open a FMIS project and after reviewing all 61 important fields, finding all fields "true", and the project funding authorization is completed with formally electronically signing of the federal electronic document. This data will support further internal analysis and discussions in an attempt to identify why the user reliability estimates (20 percent to 50 percent in the first attempt) are in such contrast and what could be expected for this important and time consuming activity. This could help streamline the process in the FHWA Wisconsin Division office. This may take coordination with the Wisconsin Department of Transportation counterparts to stress the importance of "true" information and load the FMIS fields accurately and correctly.

This work also demonstrates how important is the quality of data for successful steps in the decision on project funding authorization reviews in the FMIS software. If any one of the 61 fields is not entered correctly or missing, the project would remain pending. Therefore each employee/user should be briefed on the importance of each field of information to be entered correctly and in the format specified and recognized by all the subsequent users.

A number of strategies and methodologies from the data quality management literature (Even and Shankaranarayanan, 2007, p. 75; McGill *et al.*, 2003; Satzinger and Olfman, 1998; Fang and Neufeld, 2006; Henderson and Murray, 2005) could be considered in an effort to improve FMIS data quality dimensions of completeness, validity, accuracy, currency, and context, as well as for data quality management approaches. This could involve the use of "wrapping" (the use of embedded code to make FMIS self-verifying) and "the subsequent phased introduction of software to minimize failures" (Knaus *et al.*, 2004). Or a systematic analysis and quality management approach could be taken (Ravichandran and Rai, 1999/2000; Chow and Lui, 2003). If time and resources were available, further study could consider several aspects of this reliability analysis project. Looking at the most complete data sets available could be investigated to get better data to make the reliability estimate. This could involve other statistical analysis (Minium *et al.*, 1998; Norusis, 2003). Finally, the idea of "repairability" could be looked at as the FMIS web-based process fields which can be changed or filled-in on-line. This could allow an estimate of number of times a FMIS project has to be revisited, or how complete or true the fields are when the project is first examined. Overall, the FMIS software could be investigated from the perspective of information quality (Sen, 2001; Parssian *et al.*, 2004), to better understand and improve the data for decision makers and end users. Also, various improvement methodologies could be utilized, such as requirements negotiation (Data & Analysis Center for Software, 2008), or design considerations for web-based applications (Wroblewski and Rantanen, 2001). It is recognized that addressing the issue of reliability in FMIS – a national web-based environment for this customized software with numerous stakeholders at local, state, and national levels – will be problematic (Madu, 2005; Noy, 2003).

References

Chow, W.S. and Lui, K.H. (2003), "A structural analysis of the significance of a set of the original TQM measurement items in information systems function", *The Journal of Computer Information Systems*, Vol. 43 No. 3, pp. 81-91.

Data & Analysis Center for Software (2008), *Requirements – Change Management*, available at: www.goldpractices.com/practices/rto/index.php

Even, A. and Shankaranarayanan, G. (2007), "Utility-driven assessment of data quality", *Database for Advances in Information Systems*, Vol. 38 No. 2, pp. 75-93.

Fang, Y. and Neufeld, D.J. (2006), "The pendulum swings back – individual acceptance of re-centralized application platforms", *Database for Advances in Information Systems*, Vol. 37 Nos 2/3, pp. 33-41.

Federal Highway Administration (2000), *Evaluating the Performance of Environmental Streamlining: Development of a NEPA Baseline for Measuring Continuous Performance*, available at: www.environment.fhwa.dot.gov/strmlng/baseline/section3.asp

Henderson, I. and Murray, D. (2005), "Prioritizing and deploying data quality improvement activity", *Journal of Database Marketing & Customer Strategy Management*, Vol. 12 No. 2, pp. 113–19.

Jain, M. and Priya, K. (2005), "Software reliability issues under operational and testing constraints", *Asia-Pacific Journal of Operational Research*, Vol. 22 No. 1, pp. 33-49.

Kales, P. (1998), *Reliability for Technology, Engineering, and Management*, Prentice-Hall, Upper Saddle River, NJ.

Knaus, R., Aougab, H. and Bentahr, N. (2004), *Software Reliability: A Preliminary Handbook*, Publication FHWA-RD-04-080, US Department of Transportation, available at: www.tfhrc.gov/safety/pubs/04080/04080.pdf

McGill, T., Hobbs, V. and Klobas, J. (2003), "User-developed applications and information systems success: a test of DeLone and McLean's Model", *Information Resources Management Journal*, Vol. 16 No. 1, pp. 24-45.

Madu, C.N. (2005), "Strategic value of reliability and maintainability management", *International Journal of Quality & Reliability Management*, Vol. 22 Nos 2/3, pp. 317-28.

Minium, W., Clarke, R.B. and Coladarci, T. (1998), *Elements of Statistical Reasoning*, John Wiley & Sons, New York, NY.

Norusis, M.J. (2003), *SPSS 12.0 for Windows Student Version*, Prentice-Hall, Chicago, IL.

Noy, O. (2003), "Ensuring reliability of customized software in distributed systems", *Computer Technology Review*, Vol. 23 No. 7, pp. 45-6.

Park, N. and Sanders, C. (2007), "Improved database management yields pipeline integrity benefits", *Oil & Gas Journal*, Vol. 105 No. 22, pp. 62-5.

Parssian, A., Sarkar, S. and Jacob, V.S. (2004), "Assessing data quality for information products: impact of selection, projection, and Cartesian product", *Management Science*, Vol. 50 No. 7, pp. 967-82.

Ravichandran, T. and Rai, A. (1999/2000), "Total quality management in information systems development: key constructs and relationships", *Journal of Management Information Systems*, Vol. 1 No. 2, pp. 119-55.

Satzinger, J.W. and Olfman, L. (1998), "User interface consistency across end-user applications: the effects on mental models", *Journal of Management Information Systems*, Vol. 14 No. 4, pp. 167-92.

Sen, D. (2001), "Does the measure of information quality influence survival bias?", *International Journal of Quality & Reliability Management*, Vol. 18 Nos 8/9, pp. 967-81.

US Government Accountability Office (2005), *Accountability Office Federal-Aid Highways – FHWA Needs a Comprehensive Approach to Improving Project Oversight*, US Government Accountability Office, Washington, DC, available at: www.gao.gov/new.items/d05173.pdf

Wroblewski, L. and Rantanen, E.M. (2001), "Design considerations for web-based applications", paper presented at the 45th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA.

About the authors

Mark R. Chandler, PE, CMfgE, is a Field Operations Engineer with environmental, design and construction responsibilities in the USDOT Federal Highway Administration (FHWA), Wisconsin Division. Prior to current position with FHWA he was most recently employed at Honeywell's Solid State Electronic Center, and prior to that the US Bureau of Mines. Education includes an undergraduate degree in Geo-Engineering, a Master's in Manufacturing Systems Engineering with a Quality Concentration, and currently as a PhD student at Bowling Green State University in Technology Management with a Quality Systems Specialization. Mark R. Chandler is the corresponding author and can be contacted at: mark.chandler@fhwa.dot.gov

M. Affan Badar, PhD has been a faculty member at the Indiana State University (ISU) since Fall 2002. He has been also serving as the Assistant Director of the ISU Center of Systems Modeling and Simulation since 2004. He served the IIE Engineering Economy Division as the Director since 2005 through 2007. He received his PhD degree in Industrial Engineering from the University of Oklahoma in 2002, MS in Mechanical Engineering from King Fahd University of Petroleum and Minerals in 1993, and MSc in Industrial Engineering from the Aligarh Muslim University in 1990. At ISU he teaches courses for the BS in Mechanical Engineering Technology, MS in Industrial Technology, and PhD in Technology Management programmes. He has published more than 20 articles in refereed journals and proceedings in the areas of coordinate metrology, lean manufacturing, health care, energy system design, stochastic modeling, reliability, and supply chain cost analysis.

Applying DEA to enhance assessment capability of FMEA

Dong-Shang Chang
Department of Business Administration, National Central University, Jhongli City, Taiwan, and School of ORIE, Cornell University, Ithaca, New York, USA, and
Kuo-Lung Paul Sun
Department of Business Administration, National Central University, Jhongli City, Taiwan

Received 28 April 2008
Revised 8 November 2008
Accepted 18 January 2009

Abstract

Purpose – The purpose of this study is to propose a state-of-the-art new approach to enhance FMEA assessment capabilities.

Design/methodology/approach – Through data envelopment analysis (DEA) technique and its extension, the proposed approach evolves the current rankings for failure modes by exclusively investigating SOD in lieu of RPN and to furnish improving scales for SOD.

Findings – Through an illustrative example the proposed approach supports the proposition that DEA can not only complement traditional FMEA for improving assessment capability but also, especially, provide corrective information regarding the failure factors – severity, occurrence and detection. Further application of DEA Stratification also reveals that this methodology is useful for managing resource allocation and risk management.

Practical implications – It is shown that the proposed approach enables manager/designers to prevent system or product failures at an early stage of design. Moreover, the approach is able to provide managerial insight of SOD more effectively than justifying the efforts on RPN alone. Projection of each SOD is determined to help managers examine the scale of efforts. Finally, the stratification analysis offers the economical allocation of failure modes with respect to the incurred costs and the efficiency.

Originality/value – The paper proposes a unique new approach, robust, structured and useful in practice, for failure analysis. The methodology, within a firmed methodology, overcomes some of the largely known shortfalls of traditional FMEA: it takes into account multiple criteria and restricted weighted; and it analyses the failure modes' ranking considering not only the direct impacts of failure indices, but also the contribution of these indices.

Keywords Failure modes and effects analysis, Data analysis, Performance management, Linear programming, Decision making

Paper type Research paper

Introduction

Failure mode and effects analysis (FMEA)

Failure mode and effects analysis (FMEA) is an easy to use and yet powerful pro-active engineering quality method that helps to identify and counter weak points in the early conception phase of products and processes (FMEA Info Centre, 2005). It is a tool for engineers to show in a structural and formalized manner in subjective thinking and experiences. FMEA is a decision making tool for prioritizing corrective action to enhance product/system performance by eliminating or reducing failure rate. In FMEA there are three factors (inputs) that determine failure risk priority (output). The first input is

International Journal of Quality & Reliability Management
Vol. 26 No. 6, 2009
pp. 629-643

0265-671X
DOI 10.1108/02656710910966165

severity (S), which is the seriousness of effect of the failure. The second input is occurrence (O), which is the probability or frequency of the failure. The third input is detection (D), which is the probability of failure being detected before the impact of the effect is realized. The failure risk priority number (RPN) is a mathematical product of severity (S), occurrence (O), and detection (D). In mathematical form, $RPN = S \times O \times D$. A typical set of failure index rankings and criteria are presented in Tables I-III. (Ford Motor Company, 1988; Field and Swift, 1996; Tang and Ho, 1996; Besterfield *et al.*, 1999; Sankar and Prabhu, 2001). A failure mode being evaluated will have failure ranking "relatively" to other failure modes. The higher the RPN, the higher the chance that the mode will fail and subsequently, this mode demands higher priority for corrective action.

However, the fundamental problem of FMEA is that it attempts to quantify risk solely through RPN computation without adequately quantifying the factors that

Table I. Typical rankings of severity indices

Effect	Criteria	Rank
Hazardous	Failure is hazardous, and occurs without warning It suspends operation of the system and/or involves noncompliance with government regulations	10
Serious	Failure involves hazardous outcomes and/or noncompliance with government regulations or standards	9
Extreme	Product is inoperable with loss of primary function The system is inoperable	8
Major	Product performance is severely affected but functions The system may not operate	7
Significant	Product performance is degraded Comfort or convince function may not operate	6
Moderate	Moderate effect on product performance The product requires repair	5
Low	Small effect on product performance The product does not require repair	4
Minor	Minor effect on product or system performance	3
Very minor	Very minor effect on product or system performance	2
None	No effect	1

Source: Ford Motor Company (1988)

Table II. Typical rankings of occurrence indices

Probability of failure	Possible failure rate	Rank
Extremely high: failure almost inevitable	$\geq$1 in 2	10
Very high	1 in 3	9
Repeated failure	1 in 8	8
High	1 in 20	7
Moderately high	1 in 80	6
Moderate	1 in 400	5
Relatively low	1 in 2,000	4
Low	1 in 15,000	3
Remote	1 in 150,000	2
Nearly impossible	$\leqq$ 1 in 1,500,000	1

Source: Ford Motor Company (1988)

Effect	Criteria: likelihood of detection	Rank
Absolute uncertainty	Control does not detect a potential cause of failure or subsequent failure mode; or there is no design control	10
Very remote	Very remote chance the design control will detect a potential cause of failure or subsequent failure mode	9
Remote	Remote chance the design control will detect a potential cause of failure or subsequent failure mode	8
Very low	Very low chance the design control will detect a potential cause of failure or subsequent failure mode	7
Low	Low chance the design control will detect a potential cause of failure or subsequent failure mode	6
Moderate	Moderate chance the design control will detect a potential cause of failure or subsequent failure mode	5
Moderately high	Moderately high chance the design control will detect a potential cause of failure or subsequent failure mode	4
High	High chance the design control will detect a potential cause of failure or subsequent failure mode	3
Very high	Very high chance the design control will detect a potential cause of failure or subsequent failure mode	2
Almost certain	Design control will almost certainly detect a potential cause of failure or subsequent failure mode	1

Source: Ford Motor Company (1988)

Table III. Typical rankings of detection indices

contribute to risk (Sankar and Prabhu, 2001). Therefore the RPN alone could not provide sufficiently important information relating to factors. Although the risk management can be accomplished by examining the RPN (between 1 and 1,000) to represent increasing level of risk in 1,000 possibilities of severity-occurrence-detection combinations, it still has several shortcomings. For example, if two or more failure modes have exact RPN, one may face difficulty in selecting which failure mode demands higher priority for corrective action.

Bowles (1998) pointed out that the FMEA scales for severity and detection are only qualitative. For example, a rank 8 in severity is not twice as a rank 4 in severity to form the RPN, the ratings are treated as if they represent numeric quantities. Bowles also stressed out that the RPN number calculation erroneously oversimplified that a two-fold increase in a factor (severity, for example) can be offset by a corresponding decrease of half in another factor (Bowles, 1998).

Furthermore, the RPN scale itself has some non-intuitive statistical properties. The initial and correct observation that the scale starts at 1 and ends at 1,000 often lead to incorrect assumptions about the middle of the scales (Sankar and Prabhu, 2001). In fact there are only 120 unique RPNs, because some of RPNs are repeating. For example, RPN 120 appears 24 times from combination of SODs. Other extremes 1, 123 and 1000 appear only once. Few instances, some hypothetical numbers do not happen at all, e.g. 13, 17, 19 etc. The traditional FMEA assumes that all RPNs have same occurrence probabilities, but that is not the case concluding from above. Figure 1 shows exhaustive listing of frequency distribution of RPN set. Figure 1 demonstrates RPN probabilities neither normally nor uniformly distributed.

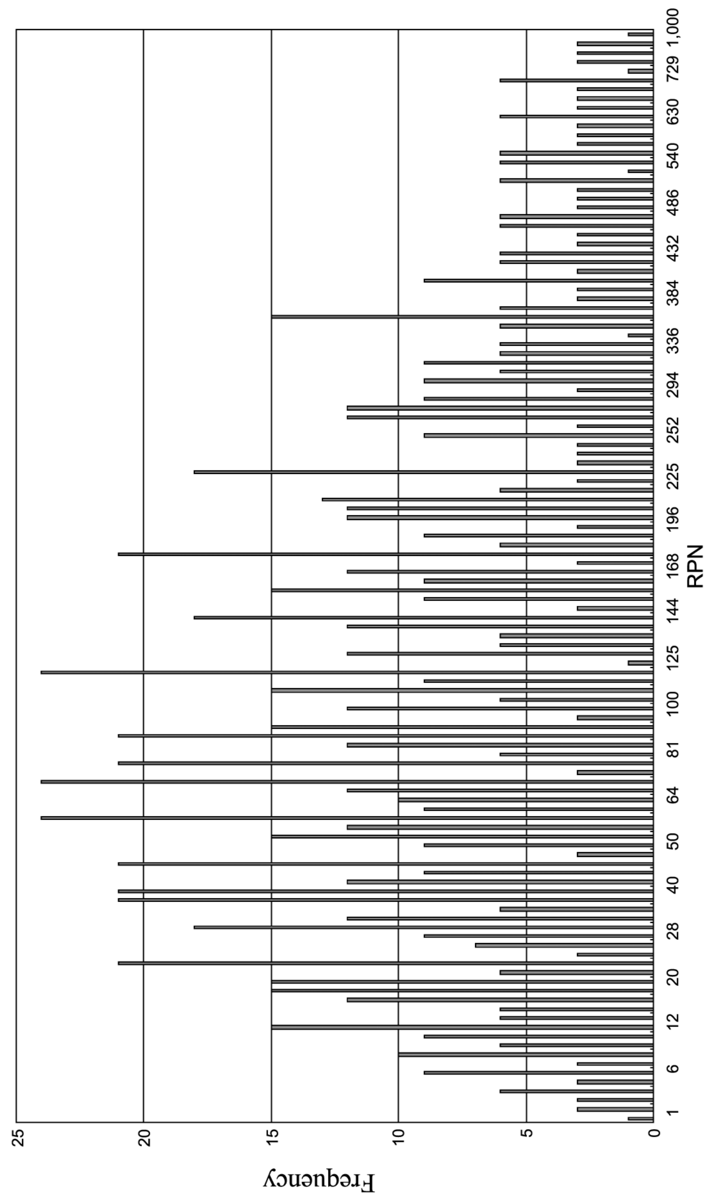

Figure 1.
Frequency distribution of RPN

There are some other considerable problems involving RPN generating. Problems such as neglect interdependencies among various failure modes and its effects; and that the same design level corresponds to the same RPN values on different index scales (the assumption is that the three factors are all equally important) (Shahin, 2004).

Nevertheless, there were also authors proposing different ways of trying to improve the effectiveness of FMEA. Sanker and Prabhu (2001) modified RPN by Risk Priority Rank (RPR) with yet another 1,000 possible severity-occurrence-detection combinations incorporating an expert's judgment. Puente *et al.* (2002) presented two alternative ways based on:

(1) structure expert knowledge in the form of qualitative decision rule whereby a risk priority category can be assigned to each cause of failure; and

(2) on fuzzy decision system which increase the continuity of the FMEA decision model, and which optimizes risk discrimination of difference of failure.

Most recent study, Garcia *et al.* (2005) developed fuzzy data envelopment analysis approach for FMEA, in which they combined uncertainties integrated with linguistic variables. Essentially a failure mode attached with greater uncertainties, will be considered critical and demand more attention.

Data envelopment analysis (DEA)

Data envelopment analysis (DEA) is a linear programming based methodology for evaluating performance among peers in terms of how "relatively" efficiently these peers' utilized inputs that transformed into outputs. DEA has also been used to supply new insights into activities and entities (Charnes *et al.*, 1978; Charnes *et al.*, 1994; Cooper *et al.*, 2004). The efficiency or performance measure of a Decision Making Unit (DMU) is defined by its position relative to the frontier of best performance established mathematically by the ratio of weighted sum of outputs to weighted sum of input (Norman and Stoker, 1991). DEA can answer what inputs and what quantity are to be reduced; what outputs and what amounts are to be increased so as to be efficient. Furthermore, SODs release failure likelihood in terms of judgment scaling against failure modes. On the other hand, DEA examines inputs and outputs to tender relative efficiency scores among DMUs.

Owing to the characteristics of DEA, it is deemed a good alternative tool to enhance the assessment capability of FMEA. The common features of DEA and FMEA are such:

- determination of risk ranking for failure modes in FMEA as to efficiency ranking of DMUs in DEA; and
- both deem to have inputs and outputs. In FMEA the inputs are severity, occurrence and detection and output is RPN.

RPNs are presented as relatively risky among failure modes. Prioritized corrective actions are tackled based on the RPN derived from the inputs. Here, the failure modes used in FMEA is equivalent to DMUs in DEA; the inputs (severity, occurrence, detection) of FMEA can be identically seen as multiple inputs of DEA. However, here in FMEA we want RPN as small as desired, whereas in DEA we prefer efficient (outputs/inputs) as great as possible.

Objective

By applying DEA in the assessment of FMEA, we define three objectives for this study. These are:

(1) to identify risk rankings among failure modes in terms of cardinal order;

(2) to provide quantitatively corrective information for risk indexes (severity, occurrence, detection) that subsequently reduce failure probability; and

(3) to imply resource allocations management based on the clustering of entire failure modes.

Methodology

This section gives rolling theoretical process of data envelopment analysis (DEA) on failure mode and effects analysis (FMEA). First part describes the principle of DEA basic model, which was proposed by Charnes, Copper, and Rhodes (CCR) in 1978. And briefly states implicit correlation between FMEA and DEA. Second, we incorporate input multipliers constraints to form CCR Assurance Region (CCR AR). Demonstrate how DEA defines ranking orders as well as weightings through application of input-oriented CCR, multiplier model and input-oriented CCR AR model. Finally, the principle of stratification of DEA is articulated. In this study and succeeding sections, the terms of failure modes in FMEA and decision-making units (DMUs) in DEA are interchangeable.

Input-oriented DEA CCR model

Given a set of decision-making units (DMUs) $j = 1, \ldots, n$ utilizing quantities of inputs $X \in x^m$to produce quantities of outputs $Y \in y^s$. Denotes x_{ij} the amount of the ith input used by the jth DMU and y_{rj} the amount of the rth output produced by the jth DMU. Then efficiency score of the DMU, θ_j is measured as:

$$\theta_j = \frac{\sum_{r=1}^{s} \mu_{rj} y_{rj}}{\sum_{i=1}^{m} v_{ij} x_{ij}} \text{ for } j = 1, \ldots, n$$

where μ_{rj} is the output weight and ν_{rj} is the input weight.

Considering a number (n) of DMUs, the efficiency score θ_j is then maximized subject to:

$$\frac{\sum_{r=1}^{s} \mu_{rj} y_{rj}}{\sum_{i=1}^{m} v_{ij} x_{ij}} \leq \text{for } j = 1, \ldots, n$$

and $\mu_{rj},\ \nu_{ij} \geq 0$

The first inequality promises that the efficiency scores of other DMUs cannot exceed one, while the second one requires the weights to be positive. The weights for each output and input are determined by DEA so that each DMU maximizes its own efficiency ratio. Any other set of weights results in a lower efficiency score. In other words, DEA gives the benefit of the doubt to each DMU in calculating the efficiency ratio. The above non-linear problem can be transformed into a linear one and be solved. For a linear programming problem, there exists a pair of expressions – primal and dual forms. The primal linear program is also called "multiplier" DEA model as follows:

$$\max z = \sum_{r=1}^{s} u_r y_{ro}$$

subject to:

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_r x_{rj} \leq 0 \text{ for } j = 1, 2, \ldots, n$$

$$\sum_{\mathrm{i}=1}^{m} v_i x_{io} = 1$$

$$u_r, v_i > \varepsilon > 0$$

And the dual form (also envelopment form) to Equation (3) is:

$$\min \theta - \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right)$$

Subject to:

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io}\ i = 1, 2, 3, \ldots, m;$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro}\ r = 1, 2, 3, \ldots, s;$$

$$\lambda_j \geq 0\ j = 1, 2, 3, \ldots, n.$$

$$\varepsilon > 0$$

where ε is non-Archimedean infinitesimal S^- and S^+ are input slacks and output slacks respectively.

Equation (4) represents a two-stage DEA process with maximal reduction of inputs being achieved first, it yields efficient scoring (θ). Then, in the second stage, movement onto the efficient frontier is achieved via optimizing the slack variables (S^- and S^+). The Equation (3) and (4) are so called "DEA CCR model".

In sum, Equation (4) will render DMUs efficiency scores and input slacks while Equation (3) returns weightings of input. A DMU is efficient if and only if the following two conditions are satisfied:

(1) $\theta^* = 1$; and

(2) all slacks are zeros.

The nonzero slacks and the value of $\theta^* \leq 1$ identify the resources and amount of any inefficiencies of failure mode that may be presented (Cooper *et al.*, 2000).

Owing to the RPN is product of SODs, the computed RPN is no longer relevant in DEA application. Therefore we set the output of each failure mode to one in subsequent DEA application. The input-oriented CCR model, in which, the objective is to minimize inputs while producing at least the given output levels. The results from DEA model provide both primal (multiplier form) and dual (envelopment form) solutions as well as slacks, and target projections onto efficient frontier. Initially, in the original input-oriented CCR model that input variables are not limited to ceiling boundary (Cooper *et al.*, 2000).

Input-oriented DEA CCR assurance region (AR) model

The input-oriented CCR model, gives no boundaries on the weightings of input variables, therefore the unrestricted constraint could render unrealistic estimations. This freedom of choice shows the DMU in the best possible light, and is equivalent to assuming that no input is more important than any others. Decision making has, in some contexts, value judgments that can be formalized a priori. These value judgments can reflect known information about how the factors used by the DMUs behave, and/or "accepted" beliefs or preferences on the relative worth of inputs or even DMUs (Thanassoilis and Portela, 2004). Considering the inputs of SOD are bounded to between 1 and 10, CCR AR incorporates paired ratio onto input multipliers reflecting unbiased efficient frontier. Introducing weight restriction is an extension of multiplier form of DEA. To illustrate the approach of AR, we incorporate additional inequality constraints of the following into the multiplier DEA models in Equation (3) as:

$$\alpha_i \leq v_i/v_{io} \leq \beta_i,\ i = 1, \ldots, m$$

Here, v_{io} represents multipliers which serve as "numeraires" in establishing the upper and the lower bounds denoted here by α_i, and β_i, for the multipliers associated with inputs $i = 1, \ldots, m$ (Cooper *et al.*, 2000; Cook and Zhu, 2005). The above additional constraints (5) attached to Equation (3) are so called "CCR Assurance Region model".

As we apply weight restriction onto Equation (3), the DEA AR model yields input target projections, which render a DMU Pareto-efficient. In the absence of weights restriction such target projections have the following two features:

(1) they preserve as far as possible its observed mix of input and output levels; and

(2) they involve no deterioration to any observed input or output level.

These two features of target projections can be altered when the targets are obtained from a DEA model with incorporating binding weights restrictions. In other words, in respect of DMU j_o a target input level may be higher (i.e. worse) than the corresponding observed input level (Thanassoulis, 2001).

Stratification of DEA on failure modes

Considering individual failure modes neighborly posts near-same risk degree to each others resulting in dilemma decision, we apply DEA Stratification approach to grouping failure modes in risk layer rather than individuals. Once the ranking orders (efficiencies) of all failure modes have been determined, we further conduct clustering for both efficient and inefficient failure modes altogether. As for a decision maker who will appreciate to know which modes stand same risk level in terms of failure probability so that resources can be dispensed more effectively. In this regard, that adding or deleting an inefficient DMU or a set of inefficient DMUs do not alter the efficiencies of the existing DMUs and the best-practice frontier (Zhu, 2003).

Thus, by repeating Equations (3) and (5) yields stratifications of whole set of DMUs. So long as no more inefficient DMUs left in last level of efficient frontier, then the clustering process is completed. Interested readers may refer to Zhu's (2003) for detailed algorithmic iteration.

To obtain the first level of stratification (first best-practice frontier), we apply CCR AR for initial run. Then remove DMUs of first level to allow the remaining inefficient DMUs to extract a new second-level best-practice frontier. If we remove this second-level best-practice frontier, a third-level best-practice frontier is extracted by re-applying CCR AR, and so on, until no DMU is left. In this manner, we partition the set of DMUs into several levels of best-practice frontier (Zhu, 2003).

An illustrative example

For demonstrative purpose, an example was taken from a study of software risk management. Dillibabu and Krishnaiah (2006) applied failure mode and effects analysis (FMEA) to software coding in an attempt to improve defect-free software products in the design and development stage. In their study, 11 failure modes were identified as well as index scorings were assigned. In the present study, we excerpted matrix of S O D and its RPN. Original FMEA data format is shown in Table IV. In the following parts, first we converted FMEA data matrix into DEA data format. Next, efficiency scores and slacks of failure modes were computed. Follow by input weights, input projections and their reduction rate. Final part, the stratification of failure modes was obtained to draw group reference base for failure modes. The entire computational processes were carried out by the DEA EXCEL SOLVER developed by Zhu (2003).

Efficiency scores

Prior to applying input-oriented CCR AR model, the restricted weights of SOD is incorporated given by bounded weight ratio of $\alpha_i \leq v_i/v_{io} \leq \beta_i$,, where $i = 1, \ldots, m$,

No.	Failure mode	S	O	D	RPN
1	Inadequate skill level of reviewers	7	4	10	280
2	Review objective not defined	3	8	5	120
3	Inadequate time	5	1	5	25
4	Inappropriate methodology	3	4	1	12
5	Review issues not recorded	10	4	5	200
6	Process non-compliance	5	8	1	40
7	Review time not utilized	1	10	1	10
8	Plan non-compliance	1	4	1	4
9	Verification of issues not complete	10	10	5	500
10	Issues not address	7	4	10	280
11	Incorrect fixes	10	1	10	100

Source: Dillibabu and Krishnaiah (2006)

Table IV. Original FMEA data

where α_i is obtained by $\min(v_i)/\max(v_{io})$ and β_i is obtained by $\max(v_i)/\min(v_{io})$ respectively. Thus their input weight ratios boundaries are determined as follows:

$$0.1 \leq O/S \leq 10.0;\ 0.1 \leq S/D \leq 10.0;\ 0.1 \leq O/D \leq 10.0$$

For the purpose of comparison, we deployed input-oriented CCR and CCR AR to see how weight constraints affect the outcomes. In the Table V, CCR returned six "efficient failure modes" whereas CCR AR yielded only two modes. Obviously, the CCR AR model has better discriminating power in risk evaluation. Furthermore, by the definition of true efficient is that any mode that is "efficient" pertaining no slack whatsoever. In this regard, even though the CCR model identified mode #4, 6, 7 and 11 having efficient score equaling to 1, they still possess some slacks. However, the CCR AR does not show such contradiction.

Per efficiency scoring assessment for FMEA implies that failure mode #9, Verification of issues not complete, ($\theta = 0.33$) has highest risk level relatively to other failure modes. From the software coding stand point, final verification is perhaps most critical issue before releasing. The next higher risk is mode #2, Review objective not define, with efficiency score 0.45. The implication is that if objective is not pre-defined it may tackle wrong effort. Table V also confirms that non-Pareto-efficient modes obtained from CCR AR model are lower than CCR. Other risk orders and its implications are apparent by examining their efficiency scores respectively.

Weighting of SOD

To confirm the justification of CCR AR model, the weightings of SOD from both CCR and CCR AR for each failure mode are computed as shown in Table VI. Clearly, CCR AR model has reasonable weightings for SOD than CCR model does. In such, CCR model resulted in many zero weights. Note that the total weighted sum of SODs in CCR AR model is determined as one (1) to reflect the relative importance in risk evaluation.

Examining failure modes #1 and #2, for example, CCR model identified Detection having null weighting with respect to their failure modes. On the other hand, CCR AR model confirmed Detection having weights of 0.10 and 0.05 for mode #1 and mode #2 respectively. By the same token, other failure modes signal similar findings between CCR model and CCR AR model.

	Results from CCR slack				Results from CCR AR slack			
DMU or failure mode	θ	S	O	D	θ	S	O	D
1	0.51	0.00	0.00	1.54	0.50	3.52	1.86	6.52
2	0.46	0.00	0.06	0.93	0.45	1.64	4.27	3.63
3	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
4	1.00	2.00	0.00	0.00	0.96	2.00	0.00	0.00
5	0.61	3.06	0.00	0.00	0.58	7.12	1.41	2.12
6	1.00	4.00	4.00	0.00	0.65	4.00	4.00	0.00
7	1.00	0.00	6.00	0.00	0.80	0.00	6.00	0.00
8	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
9	0.35	1.73	0.00	0.00	0.33	8.33	6.50	3.33
10	0.51	0.00	0.00	1.54	0.50	3.52	1.86	6.52
11	1.00	5.00	0.00	5.00	0.67	5.00	0.00	5.00

Table V.
Efficiency scores and slacks

DMU or failure mode	Results from CCR S	O	D	Results from CCR AR S	O	D
1	0.08	0.11	0.00	0.48	0.42	0.10
2	0.07	0.10	0.00	0.19	0.76	0.05
3	0.00	0.21	0.16	0.68	0.21	0.11
4	0.00	0.21	0.16	0.06s	0.75	0.19
5	0.00	0.13	0.10	0.12	0.48	0.40
6	0.00	0.00	1.00	0.22	0.35	0.43
7	0.00	0.00	1.00	0.33	0.33	0.34
8	0.16	0.21	0.00	0.20	0.78	0.02
9	0.00	0.07	0.05	0.07	0.70	0.23
10	0.08	0.11	0.00	0.48	0.42	0.10
11	0.00	1.00	0.00	0.33	0.33	0.34

Table VI. Weightings of SOD

Moreover, by applying weight ratio restriction to multiplier, the sum of weighting of SOD equals to 1. Result of CCR model in Table V, mode #6 for example, the efficiency score is 1 and its SOD slacks are 4.00, 4.00 and 0.00 respectively. And its weights of SOD in Table VI are 0.00, 0.00 and 1.00 respectively; these zeros lead to mis-assessment (because S and O contributed nothing). Whereas the CCR AR in Table V, mode #6 posts efficiency score of 0.65; and in Table VI its weights of 0.22, 0.35 and 0.43 for SOD respectively. From the preceding discussion and CCR AR confirmation, this study concludes that CCR AR approach rationally serves best justification.

Input projections and reduction rate

The proposed inputs projections of SODs could render the inefficient DMUs into the efficient frontier. The inputs projections are straightforwardly obtained from CCR AR model. In the Table VII, for those modes or DMUs whose efficiency are less than one may have input projections (new input targets) against their original value of SOD. Subsequently, projection values and reduction rates of each SOD are easily calculated as shown in the two last columns.

Stratification of failure modes

By applying both Equation (3) and (5) once at a time, so long as all of DMUs have situated their level of efficient frontier, then the stratifying process is completed. In the Table VIII, cluster classification and its member(s) are presented. The first iteration of CCR AR returned mode #3 and #8 as first level. This first level demands least corrective action (or the least risk endurance) than the subsequent levels. We then omitted these two modes and subsequently obtained level 2 and so on. There was a special characteristic that all modes in the same level tag efficient ($\theta = 1$). Thereby, we were able to classify 4 efficient levels and all failure modes exhaustedly falling into, at most, one level.

Discussion

Reviewing Table V that ranking order for improving is 9, 2, 1, 10, 5, 6, 11, 7, 4, 3, 8; and from RPN ranking perspective, the ranking is 9, 1, 10, 5, 2, 11, 6, 3, 4, 7, 8. The findings of the case study highlighted the potential gaps between input criteria (SOD) oriented

Table VII.
Input projections and reduction rates

Mode/DMU	Index	Origin	Projection	Reduction (%)
1	S	7	3.48	50.26
	O	4	2.14	46.53
	D	10	3.48	65.18
2	S	3	1.36	54.55
	O	8	3.73	53.41
	D	5	1.36	72.73
3	S	5	5.00	0.00
	O	1	1.00	0.00
	D	5	5.00	0.00
4	S	3	1.00	66.67
	O	4	4.00	0.00
	D	1	1.00	0.00
5	S	10	2.88	71.21
	O	4	2.59	35.23
	D	5	2.88	42.42
6	S	5	1.00	80.00
	O	8	4.00	50.00
	D	1	1.00	0.00
7	S	1	1.00	0.00
	O	10	4.00	60.00
	D	1	1.00	0.00
8	S	1	1.00	0.00
	O	4	4.00	0.00
	D	1	1.00	0.00
9	S	10	1.67	83.33
	O	10	3.50	65.00
	D	5	1.67	66.67
10	S	7	3.48	50.26
	O	4	2.14	46.53
	D	10	3.48	65.18
11	S	10	5.00	50.00
	O	1	1.00	0.00
	D	10	5.00	50.00

Table VIII.
Cluster classification and its member

Cluster (level)	Members (mode)
1	3 – Inadequate time
	8 – Plan non-compliance
2	4 – Inappropriate methodology
	7 – Review time not utilized
	11 – Incorrect fixes
3	1 – Inadequate skill level of reviewers
	2 – Review objective not defined
	5 – Review issues not recorded
	6 – Process non-compliance
	10 – Issues not addressed
4	9 – Verification of issues not complete

and RPN oriented ranking methodology. However the proposed approach revealed more information about SOD and furnished with suggested improving direction.

In Table VIII that clustered levels are $L_4 = \{9\}$, $L_3 = \{1, 2, 5, 6, 10\}$, $L_2 = \{4, 7, 11\}$ and $L_1 = \{3, 8\}$. The stratification DEA provides more insight for decision maker. Suppose, each mode in the higher level represents an option, decision maker may compare modes that are currently in the same level as well as compare with relevant modes in the lower level serve as evaluation context. For example, comparing mode #1 in level 3 and mode #9 in level 4 respectively, intuitively we may resolve mode #9 as first priority. However, there could be a trade-off alternative. Detection of mode #1, for example, presumably renders an easier resolution than the detection of mode #9. Thereby, the clustering is very helpful to management when decision maker faces another choice of same risk level but in various efforts on criteria. It is common to tackle for easier and yet effective modes so as to lessen failure more effectively.

In the case of more than one failure modes having same RPN indexes, decision maker would run into another decision making. DEA scoring, projection, and clustering provide more information than that of traditional FMEA does. If the economic cost of reducing severity, reducing occurrence and increasing detection is known, one can simply convert this economic cost with help of target projection inputs. It is easier to calculate the efforts with respect to the resources needed among those having same RPN indices. For example, mode #9 requires 83.33 per cent (S), 65.00 per cent (O) and 66.67 per cent (D) of reduction ratio respectively; and that reduction ratio will be the cost/risk ratio base. In addition, the benchmarking modes (efficient modes) are another good assistant tool for decision making. Managements always have context-based to making a comparison once the efforts and the resources have been dispensed on those high-risk modes.

Conclusions and suggested future research direction

Although one can rank the priority for corrective action from traditional RPN, there are more advantages that DEA can provide and yet easy deployment. The proposed approach is able to enhance the assessment capability of FMEA and to obtain managerial implication with particularly regard to corrective actions. First, so long as severity, occurrence and detection are provided, DEA can determine risk ranking with or without computing RPN at first. This technique is done by setting up the RPN at unity while directly computing SOD with DEA logarithmic input oriented assurance region model. The DEA returns rankings for all failure modes by means of computed efficiencies. Second, DEA is able to quantifiably identify the amount of severity, occurrence and detection that needs adjustment on ordering to bringing failure mode efficient or, in other words, to lessen risk. Third, the inefficient failure modes always have reference mode for benchmarking. The advantage of this benchmarking is to pursue economical allocation of resources for management decision-making. Furthermore, modes that fall into same risk group demand equal attention.

In our illustrative example, the input scales (from 1 to 10) are standard in nature. As for future study, further employs the input oriented bounded variable DEA model in Equation (4) and (5) to deal with situation of inputs ratings that exceed 10. Other alternative DEA model such as additive model and the allocation model (cost) are worth investigating. In practical world cases, if one or more of SODs cannot be altered; or some interactions occurred within SODs, it posts interesting for further studying. Some FMEA cases other than software setting can be further investigated, for instance, special applications on the patient safety in healthcare industry.

References

Besterfield, D.H., Besterfield, V.M., Besterfield, G.H. and Besterfield, M.S. (1999), *Total Quality Management*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.

Bowles, J.B. (1998), "The new SAE FMECA standard", *Proceedings of the Annual Reliability and Maintainability Symposium, Anaheim, CA, 19-22 January*, pp. 48-53.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), "Measuring the efficiency of decision-making units", *European Journal of Operation Research*, Vol. 2, pp. 429-44.

Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. (1994), *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers, Norwall, MA.

Cook, W.D. and Zhu, J. (2005), *Modeling Performance Measurement: Applications and Implementation Issues in DEA*, Springer, New York, NY.

Cooper, W.W., Seiford, L.W. and Tone, K. (2000), *Data Envelopment Analysis: A Comprehensive Text with Model, Applications, References and DEA-Solver Software*, Kluwer Academic Publishers, Norwall, MA.

Cooper, W.W., Seiford, L.M. and Zhu, J. (2004), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Norwall, MA.

Dillibabu, R. and Krishnaiah, K. (2006), "Application of failure mode and effects analysis to software code reviews – a case study", *Software Quality Professional*, Vol. 8 No. 2.

Field, S. and Swift, K.G. (1996), *Effecting a Quality Change: An Engineering Approach*, Arnold, London.

FMEA Info Centre (2005), available at: www.fmeainfocentre.com

Ford Motor Company (1988), *Potential Failure Mode and Effects Analysis (FMEA) Reference Manual*, Ford Motor Company, Detroit, MI.

Garcia, P.A.A., Schirru, R., Frutuoso, E. and Melo, P.F. (2005), "A fuzzy data envelopment analysis approach for FMEA", *Progress in Nuclear Energy*, Vol. 46, pp. 359-73.

Norman, M. and Stoker, B. (1991), *Data Envelopment Analysis: The Assessment of Performance*, John Wiley & Sons, New York, NY.

Puente, J., Pino, R., Priore, P. and de la Fuente, D. (2002), "A decision support system for applying failure mode and effects analysis", *International Journal of Quality & Reliability Management*, Vol. 19, pp. 137-50.

Sankar, N.R. and Prabhu, B.S. (2001), "Modified approach for prioritization of failure in a system failure mode and effects analysis", *International Journal of Quality & Reliability Management*, Vol. 18, pp. 324-35.

Shahin, A. (2004), "Integration of FMEA and the Kano model: an exploratory examination", *International Journal of Quality & Reliability Management*, Vol. 21, pp. 731-44.

Tang, S.H. and Ho, S.Y. (1996), "Failure mode and effects analysis: an integrated approach for product design and process control", *International Journal of Quality & Reliability Management*, Vol. 13, pp. 8-26.

Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis*, Kluwer Academic Publishers, Norwall, MA.

Thanassoilis, E. and Portela, M.C. (2004), "Incorporating value judgments in DEA", in Cooper, M., Seiford, L.M. and Zhu, J. (Eds), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Norwall, MA.

Zhu, J. (2003), *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*, Kluwer Academic Publishers, Norwall, MA.

Further reading

Zhu, J. (1996), "DEA/AR analysis of the 1988-1989 performance of the Nanjing Textile Corporation", *Annals of Operation Research*, Vol. 66, pp. 311-35.

About the authors

Dong-Shang Chang is a Professor in the Department of Business Administration, National Central University, Taiwan.

Kuo-Lung Sun is a PhD candidate in the Department of Business Administration, National Central University, Taiwan. Kuo-Lung Sun is the corresponding author and can be contacted at: psun98@yahoo.com