From evaluation to meta-evaluation of engineers' training in the automotive industry

Brittle, Robert J.

From Evaluation to Meta-Evaluation of Engineers' Training in the Automotive Industry

Robert J. Brittle

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the School of Education.

University of Durham

1999

Dedicated to my father

John Brittle

September 7th 1935 - December 27th 1998.

The mind of man is this world's true dimension,

And knowledge is the measure of the mind;

And as the mind in her vast comprehension

Contains more worlds than all the world can find,

So Knowledge doth itself far more extend

Than all the minds of men can comprehend.

A Treaty of Human Learning
1st Baron Brooke (1633)

Abstract

Name: Robert J. Brittle

Title of Project: 'From Evaluation to Meta-Evaluation of Engineers' Training in the Automotive Industry'

This thesis presents and analyses the evaluation of an European wide training programme aimed at engineers working in a large multinational automotive company. The training programme is unique in that it was conceived to address particular operational concerns and involved a multicultural workforce from six European countries.

The evaluation of the training, which extends from the pilot stages of the programme through to its full implementation, where Kirkpatrick's four level evaluation framework is used, is the company's first large scale attempt at systematic training evaluation. The evaluation of the programme is typical in its approach as reflected in the wide body of literature, however the use of meta-evaluation to determine the overall value of the evaluation approach in a commercial context provides originality and the basis for establishing an alternative approach to evaluating vocational training.

The main body of the thesis is presented in three parts. Part I provides a critical review of the literature relating to; learning and training; conceptualisations of evaluation; and measurement and evaluation methodology, to establish the foundation for the empirical study. Part II is a detailed analysis of the evaluand, the evaluation methodology employed, and the results and outcomes from the evaluation. Part III provides directions for training evaluation based on a meta-evaluation of the empirical study.

The thesis draws conclusions with respect to the role of evaluation in organisational training. The evaluation of training is largely conceptualised in the literature as being concerned with the <u>assessment of value or worth</u> of training to an organisation, which is the prevailing paradigm of Kirkpatrick's training evaluation framework. From the evidence obtained through the empirical study with regard to utility, feasibility, propriety and accuracy, it is concluded that the role of evaluation should be directed towards <u>maximising value or worth</u> of training through the systematic assessment, feedback and optimisation of the identifiable parameters of the training process, with the outcomes of training forming part of an overall evaluation of training framework.

Acknowledgements

I should like to acknowledge the guidance and assistance given to me in this research by my supervisors at the University of Durham; to Jack Gilliland whose encouragement and pragmatism gave me the confidence to undertake a project of this scale and whom I miss since his retirement; and to John McGuiness whose guidance and counsel not only assisted me in my studies, but helped me through a very difficult time in my life.

I should also like thank all my work colleagues who have given their time, and in particular Dr Ed Henshall whose sponsorship enabled me to start this work and whose continued support is much appreciated and Andy Bern for his understanding and for providing me with the opportunities to attend to this work.

Finally, I should like to acknowledge the support from my family; from James and Melissa for their understanding and patience during the long periods of time when we should have been doing things together; from my partner, Kelly; from my parents, Patricia and John; and my brother and sister-in-law, Peter and Carole.

Table of Contents

Appendices

List of Figures

List of Tables

Introduction

Introduction

The evaluation of training has been described in the literature as an illusive activity (e.g. Goldstein et al, 1989; Patrick, 1992) which has not been integrated into the theory and practice of industrial and commercial training. Campbell (1989), in providing an agenda for training research, asserts that training evaluation should go beyond answering whether training interventions work in terms of their outcomes by exploring how and why learning and performance occurs. In this sense evaluation cannot be treated in isolation to the theory and practice of learning and training.

Despite the volume of literature on the subject, there is little evidence to suggest that training evaluation has successfully been put into practice. French (1953) reported that only one company in forty made any attempt to evaluate supervisory training programmes. The Training Agency (1989) reported that four out of every five establishments in Great Britain make no attempt to evaluate training. Patrick (1992) comments that training is evaluated too infrequently and the little evaluation which is done is mainly confined to the measurement of trainees' reactions. Even very recently, Brown (1998) in a review of corporate training evaluation concluded that evaluation of any kind is not universally done, even though there a need to justify training and to improve our practice of training.

The current state of training evaluation can therefore be described as problematic in that; (i) it's treatment in the literature is largely independent of the bodies of knowledge pertaining to learning and training; and (ii) theory has not been widely put into practice. Given the current state of training evaluation, particularly in these two respects, the investigation that lies at the heart of this research has led the researcher into a complex web of intertwined problems and sub-problems.

From the reviews of the literature presented in part I, through the treatment of the empirical study in part II, to the analysis and conclusions in part III, the researcher has attempted to draw on and embrace the broad range of relevant concepts and elaborate the issues that emerge. It is important, therefore, at the outset to articulate clearly the specific focus of the study whilst maintaining the critical broad spectrum element to the research.

Purpose of Study

In 1989, a large automotive company initiated a pan-European training programme for all its technical personnel and the programme was implemented in six European countries. It continued for more than six years before transitioning into a global technical training programme.

This programme provided an opportunity to apply and study training evaluation within its contemporary paradigm with the purpose of providing a directional framework for training evaluation within commercial organisations.

Extensive efforts have been made to evaluate the training and this study comprises those efforts and results and contrasts the utility, feasibility, propriety and accuracy of the evaluation against the company's business operations. In this sense, this study is not simply an evaluation of training; by examining the evaluation itself in the form of a meta-evaluation (chapter 7) the researcher seeks to draw conclusions with respect to how training is and could be evaluated in the given context.

The evaluation of the training was undertaken in two stages; evaluation of the pilot programme as part of the programme's development; and evaluation of the programme's implementation. The implementation stage employed Kirkpatrick's framework for training evaluation for the purpose of measuring the programme's effectiveness in terms of its stated aims.

Kirkpatrick's framework is widely regarded as the predominant framework for the evaluation of training (Shennan and Lockhead, 1996) to the extent where it has 'significantly shaped the human resource development profession' (Gordon, 1997) and has become part of the language in many large commercial organisations (e.g. Basarab and Root, 1994). The limitations and narrowness of Kirkpatrick with respect to the management of training and organisational development are explored through this study.

The training programme and its evaluation are considered in the context of the relevant literature relating to learning and training; the broader conceptualisation of evaluation; and measurement and evaluation methodologies. This literature also has relevance to the meta-evaluation and to the foundations for a new approach to the evaluation of training.

Background and Context to the Programme

In 1989, I was transferred to the company's automotive product development division in Europe and my role was one of providing training in quality related methods to the division's technical community. My position was new and had been sponsored by senior company management following an European management quality focus meeting called by the chairman of the company's automotive operations in Europe. The purpose of the meeting was to identify the inhibitors to the achievement of immediate quality improvement which, at the time, was assessed using customer reported "things-gone-wrong" and warranty repairs indices. The improvement was recognised as being necessary for the company to become a best in class producer in pursuit of its corporate mission (Figure 0-I).

In the early eighties, the company's senior management had been greatly influenced by Dr W. Edwards Deming, a leading management consultant. It was as a result of this influence and new emphasis on quality that the senior managers in Europe were called together by the chairman to address the strategic quality issues.

Vision
To be a low-cost producer of the highest quality products and services which provide the best customer value.

Mission
The company is a world-wide leader in automotive and automotive related products and services as well as in newer industries such as communications and financial services. Our mission is to improve continually our products and services to meet our customers' needs allowing us to prosper as a business and to provide a reasonable return for our stockholders, the owners of our business.

Values
How we accomplish our mission is as important as the mission itself. Fundamental to success for the Company are these basic values:
People. Our people are the source of our strength. They provide our corporate intelligence and determine our reputation and vitality. Involvement and teamwork are our core human values.
Products. Our products are the end result of our efforts, and they should be the best in serving customers world-wide. As our products are viewed, so are we viewed.
Profits. Profits are the ultimate measure of how efficiently we provide customers with the products for their needs.

Guiding Principles
Quality comes first. To achieve customer satisfaction, the quality of our products and services must be our number one priority.
Customers are the focus of everything we do. Our work must be done with our customers in mind, providing better products and services than our competition.
Continuous improvement is essential to our success. We must strive for excellence in everything that we do; in our products, in their safety and value - and in our services, our human relations, our competitiveness, and our profitability.
Employee involvement is our way of life - we are a team. We must treat each other with trust and respect.
Dealers and suppliers are our partners. The company must maintain mutually beneficial relationships with dealers, suppliers, and our other business associates.
Integrity is never compromised. The conduct of our Company world-wide must be pursued in a manner that is socially responsible and commands respect for its integrity and for its positive contributions to society. Our doors are open to men and women alike without discrimination and without regard to ethnic origin or personal beliefs.

Figure 0-I: Company Vision, Mission and Guiding Principles (1982)

Unlike previous foci on quality, Deming's involvement with the company had facilitated a shift in thinking from a 'high quality equates to high cost' paradigm to one of 'high quality equates to low cost' (Deming, 1982). Whereas before quality had been addressed by increasing cost, senior management had started to realise that improving quality would reduce overall costs if it was approached correctly. The subsequent engineers' quality improvement training programme was set in this new paradigm and was intended to shape engineers' thinking and provide them with the tools to do this.

Emphasis on Quality

The emphasis on quality had previously resulted in extensive training in quality methods in the company. This was reflected in the industry in general, and companies had tended to follow trends for a particular quality method. Following the consultation work by Deming with the company in the early eighties (Deming, 1983), the company's quality policy letter was published (Figure 0-II) which established the concept of 'Company Total Quality Excellence'. This was important to the training programme as it was to shape the philosophy of the programme and demonstrate that senior management, at least, recognised the significance of the programme's philosophy.

Policy Letter A - 5 Subject: Company Total Quality Excellence

This Policy Letter sets forth the concept of "Company Total Quality Excellence" - a concept that emphasises the importance of quality in everything that we do. Such emphasis is in keeping with our "Company Mission, Values and Guiding Principles", which states in part: "Quality comes first. To achieve customer satisfaction, the quality of our products and services must be our number one priority."

The fundamental precepts of Company Total Quality Excellence are:
Quality is defined by the customer; the customer wants products and services that, throughout their life, meet his or her needs and expectations at a cost that represents value.
Quality Excellence can best be achieved by preventing problems rather than by detecting and correcting them after they occur.
All work that is done by company employees, suppliers, and dealers is part of a process that creates a product or service for the customer. Each person can influence some part of that process and, therefore, affect the quality of its output and ultimate customer satisfaction with our products and services.
Sustained quality excellence requires continuous product and process improvement. This means, regardless of how good present performance may be, it can become better.
People provide the intelligence and generate the actions that are necessary to realize these improvements.
Each employee is a customer for work done by employees or suppliers, with a right to expect good work from others and an obligation to contribute work of high calibre to those who, in turn, are his or her customers.

The goal of Company Total Quality Excellence is to achieve superior external and internal customer satisfaction levels. Each employees commitment to the precepts of Total Quality Excellence and management's further commitment to implementation of supporting managerial and operating systems is essential to realising this goal.

The President and Chief Operating Officer is responsible for interpreting and implementing this Policy Letter. The Company's subsidiaries and affiliated companies are encouraged to adopt a similar policy.

Figure 0-II: Company Quality Policy Letter A - 5: Company Total Quality Excellence (1984)

The notion of training in quality methods was not new to the company and significant amounts of training had been previously undertaken for a range of quality related topics. Particular emphasis had been given to statistical process control (SPC) training because of its relevance to manufacturing processes and training in other quality methods had also been undertaken, although to a lesser extent than that of SPC. In 1989, the company's UK operations had been awarded the British government's Industrial Training Award in recognition of its extensive efforts in the field.

Statistical Process Control

Within the company's manufacturing operations extensive training had been undertaken on Statistical Process Control (SPC). A so-called downstream quality method (see chapter 4), it is employed for the control of the manufacturing processes. Within this context and when applied correctly, SPC improves product conformity (improving quality) and causes a direct reduction in warranty claims and customer complaints (Oakland, 1986).

The successes within manufacturing and the desire to improve quality in other areas of the company had resulted in attempts to apply the method to control and improve other non-manufacturing processes. This had met with limited success (Henshall, 1989) and had highlighted the need for a more comprehensive range of quality methods appropriate to different contexts.

Independent Quality Methods

In addition to SPC, there were also a plethora of other quality methods including Failure Mode and Effects Analysis (FMC, 1984), Team Oriented Problem Solving (FMC and Sporting Body Mind, 1986), Taguchi's Quality Engineering (Taguchi, 1986), and Quality Function Deployment (American Supplier Institute, 1987) which were the subject of internal training programmes. Many of the techniques had been promoted in virtual isolation of the other techniques and promoted by its internal sponsor as 'the latest wonder technique' (Henshall, 1989; 1990).

Teamwork and Quality Improvement

There was also a growing recognition of the impact of teamwork and other behavioural (people) skills on the performance and improvement of the company's product development and manufacturing processes.

The corporate problem solving process and subsequent training had been devised with a recognition of the importance of teamwork (FMC, 1986) and was aptly called Team Oriented Problem Solving. This integrated teamwork into the problem solving process. Feedback from participants of the UK training (Bern, 1988) was very positive concerning the inclusion of team-building elements as part of the process and this was to later shape the design of the engineers' quality improvement training programme.

Focus on People

Of the inhibitors identified by the European management group (FMC, 1987), several held an affinity to the theme "Maximising workforce competence". One of the eight working groups which were established at the meeting was given the task of addressing the inhibitors under this theme heading and identifying actions to overcome these inhibitors.

At the follow-up meeting some three months later (December 1987), the Maximising Workforce Competence Group made the following recommendation:

> *"Industrial Relations [Personnel Department] to develop comprehensive approach to ensure that training throughout the European Automotive Operations is planned and co-ordinated; that it provides proper quality focus; and is both adequate and relevant to jobs and career development."*
>
> *In-company report (1987)*

The recommendation was accepted and agreed and a training manager for the product development division in Europe was appointed in 1989.

European Education & Training Organisation

At that time, the company's automotive operations were divided into three divisions; product development group, manufacturing operations, and sales and marketing. By the nature of their activities, the product development and manufacturing operations were closely related in terms of the engineering process. Organisationally however, these two divisions were quite distinct from each other. This impacted the organisation of the company's education and training activities.

Within these two divisions, the formalised education and training of the workforce was managed separately. The manufacturing operations division, which had facilities in Belgium, France, Germany, Portugal, Spain and the UK, had an established European training manager responsible for co-ordinating the division's training needs analysis, training development and training delivery for its European-wide operations.

The product development division, situated at two sites in Europe; Koln-Merkenich, (West) Germany; and Laindon, Essex, UK, did not have a European training manager. Each location had a training supervisor managed by the site personnel activity.

The two product development training functions operated separately from each other, with only informal lines of communication between the two. As a consequence, the division's training was not sufficiently well co-ordinated between the two sites to adequately support the core business of the division - to design and develop vehicles for the European market.

European Quality Education and Training Committee

Another outcome of the senior management focus group meeting, was the establishment of a European Quality Education and Training Committee. This reported to the company's European Quality and Strategy Committee and was co-chaired by the product development group and manufacturing operations Industrial Relations directors.

The training managers for the manufacturing operations and the product development group, who reported directly to their respective Industrial Relations directors, were appointed secretaries to the committee. The remaining membership of the committee comprised of line management and quality staffs management.

Origin of the Programme

I was appointed in April 1989, reporting directly to the product development training manager and assigned with the responsibility of designing, developing and implementing a 'Quality Tool-kit' training programme for the division's c.3500 European engineers.

A provisional programme outline was conceived (Figure 0-III). This reflected the need to transfer knowledge and skills which were the subject of training to the engineering process. The content of the training was selected on the basis of my own knowledge and that of the division's Statistical Methods Office of contemporary engineering quality methods and were congruent with the European manager group's objectives.

A critical feature of the design was the supported application of quality methods in the workplace. Knowledge and skills in terms of organisational training are enablers; that is to say they in themselves are not the goal of organisational training. The goal of organisational training is to bring about change through the facilitation of learning (Walker, 1992).

Tichy and Sherman (1993) cite improved productivity and global competition as drivers of organisational change; and Deming (1982) cites the decline of Western industry. Regardless of the stimulus for change, changes are introduced to an organisation through its technology, processes, and/or people and training, therefore, can be conceptualised in two ways with respect to organisational change; (i) as training to support change; and (ii) as training to initiate change.

The former is concerned with training which is intended to support the introduction of new equipment or facilities (technology) or changes to working practices (processes). The role of training is to provide knowledge and skills to enable people to operate the new equipment or working practices. The latter is where training itself is the stimulus for change. The working practices only change as a result of the training intervention.

The former view of training is consistent with Schein (1985) who distinguishes between indirect and direct methods for organisational change. According to Schein, indirect methods are strategies such as training, reward incentives and employee appraisal which support direct methods of change such as modifying the organisation structure and work design. The latter view of training is consistent with Woodcock and Francis (1990) who argue that the development of people through training and communication are primary to bringing about change. The training programme subject of this study falls into this second category.

Module 1

WEEK	SUBJECT AREA	RESOURCE	TR'NG GROUP SIZE	
1	Company Total Quality Excellence - Context of tools and approaches e.g. TOPS, CTC, DFA, CDRS, etc.	Internal	25	x2
	Quality Data Bank - inc. Meas. & Info. Systems	Internal	25	x2
	Statistical Methods Awareness	PD/Mfg/External	25	x2
2	Quality Function Deployment (QFD)	External	25	x2
	Failure Mode & Effects Analysis (FMEA)	Internal / External	25	x2

6 Weeks ↓ QFD & FMEA APPLICATIONS - SUPPORTED IN THE WORKPLACE

Module 2

WEEK	SUBJECT AREA	RESOURCE	TR'NG GROUP SIZE	
3	Review of QFD and FMEA projects			
	Design Verification Progress Report (DVPR) & Sign-off Report (SOR)	Internal / External	25	x2
	Design for Assembly (DFA)	External	25	x2
4	'Statistical Thinking' for Engineers (inc. Exploratory Data Analysis, Engineering Statistics / SPC)	External	12 - 13	x4

6 Weeks ↓ DFA APPLICATION - SUPPORTED IN THE WORKPLACE

Module 3

WEEK	SUBJECT AREA	RESOURCE	TR'NG GROUP SIZE	
5	Review of DFA project			
	Taguchi Methods / Quality Engineering	External /	25	x2
	Design of Experiments	Internal	12 - 13	x4
6	Team Oriented Problem Solving (TOPS/8D)	Internal / External	12 - 13	x4
	Team Building	Internal / External	12 - 13	x4

6 Weeks ↓ DOE / TAGUCHI APPLICATION - SUPPORTED IN THE WORKPLACE

1 day	PROGRAMME REVIEW & INDIVIDUAL DEVELOPMENT

Figure 0-III: Proposed Engineer Quality Education & Training Program - Component Engineer Model (July'89)

To facilitate the development of the programme and its company-wide implementation, the initial design was subjected to the views of my colleagues across the organisation. This initiated a series of discussions and planning meetings from which the engineering quality improvement training programme was conceived. The programme is described in some detail in chapter 4 and whilst it developed significantly from the original design, it retained the critical features of; (i) a quality philosophy which was congruent with company's corporate mission; (ii) including engineering process related quality improvement methods; and (iii) supported application of quality methods in the workplace.

Part I

Reviews of the Literature

Reviews of the Literature

> *'It takes a great deal of history to produce a little literature.'*
>
> *Henry James(1879)*

Before I summarise the product of my efforts with my reviews of the literature, I want to share with the reader a thought I recorded (Brittle, 1995) early in my research studies:

> *"My current perception of 'literature searching' is best explained using the analogy of searching a darkened house with a torch which has a narrow light beam. The information or knowledge exists in the darkened rooms of the house and my literature search process is one of directing the torch around the room to see what is there. I cannot see the whole picture at once and I do not know if I am in the right room until I have looked at what's in it. Furthermore, I do not know how big the house is."*

A wealth of knowledge relevant to my project exists in the literature. The literature is arranged in topic areas, or fields of study - education, psychology, management studies. At the outset, my knowledge was very limited and I was unable to see the extent of what is known.

It is conceivable, to the naive researcher, that concepts and experiences in other fields of study (e.g. engineering, mathematics, and business studies) - areas other than education, psychology, and management, hold key pieces of knowledge, albeit with different labels and contexts. One could spend a lifetime simply reading and understanding and translating the literature. Eventually one has to accept, as the writer has, that considered boundaries must be set around the fields of research literature and language if a project of this kind is ever to be completed. I hope that I have set the boundaries intelligently.

Overview

Training evaluation features widely in the popular training literature and yet very relatively little has been put into practice and published. Philips (1990) observed that only 24% of 3100 executives reported that any attempted measurement of change in job related behaviour as a result of training was made. Goldstein (1991) cites a 1986 survey of top US companies to illustrate that although many used end of course 'happy sheets', few organisations performed any detailed evaluation.

As a training co-ordinator for a large automotive company, my initial thoughts about evaluation were to do with training. Of course, training is a very narrow context for evaluation, and whilst the evaluation of training (training-evaluation) is the focus of this thesis, I have found it necessary to explore the multiple dimensions of evaluation, particularly within the social contexts.

In order to begin to understand evaluation, at its most general level, evaluation can be described as a process of inquiry and conclusion. The evaluation process can be short and simple, such as the intuitive or sub-conscious part of human nature. An example of this is a quick review by a car driver of his journey to the supermarket. The traffic conditions observed and analysed on the outward journey may lead to a decision to return home via a different route.

On the other hand, evaluation can be more systematic and planned, such as a study of a manufacturing process to establish the performance of the process and how it can be improved, or a civil engineer undertaking a structural survey of a property for a would-be buyer. Similarly, an evaluation can be an in-depth inquiry into the impact and effect of a Juvenile Drugs Awareness Programme. These evaluations are considered and planned, requiring thought about the process of evaluation. Morrison (1993) offers a useful account of the generality of evaluation and concludes that it is easy to see evaluation as 'part and parcel' of every day activities.

Whether evaluation is spontaneous and casual, or considered and planned, information is collected and conclusions are reached. These conclusions may, or may not, subsequently lead to some further action. In any case, through the process of evaluation, knowledge about the phenomenon under study is gained by those involved in the enquiry process, including the evaluator, the participants in the evaluation, and the recipients of the information. A second piece of learning that occurs is acquiring knowledge about the process of evaluation. For the writer, this learning is as important, if not more so, than the former learning as this knowledge can be transferred to future situations thereby further developing our ability to learn.

Arguably, any study which aims to come to some conclusion, or to provide information to include in a decision making process is evaluation. Guba and Lincoln (1989) describe evaluation as part of a set of activities which includes research and policy studies. Collectively these are labelled 'Inquiry'.

Defining evaluation is a challenging task because, by its nature, evaluation applies to a broad spectrum of natural and social activity. Morrison (1993) describes evaluation as suffering from "the untidiness of definition - it is a catch all term which embraces vastly different activities" (p.1). He also suggests that a definition of evaluation which encompasses all evaluation activities in all contexts is likely to be too generic to be of value in specific applications, whilst a definition which is specific to a given context is likely to be too narrow to encompass the full scope of potential evaluands.

Before I settle on a definition of evaluation for this thesis, it is necessary to consider the theory and practice of evaluation and the significance of its context by developing an understanding of (education and) training within the commercial sector of society.

Training-evaluation, like any other form of evaluation is context bound. Its definition and conceptualisation is dependent on; its purpose; its context; and its methodology, which, in turn, are respectively dependent on; the role of evaluation; the evaluand and its defining scope; and the available metrology (or measurement technology). To research and develop training evaluation, and particularly its conceptualisation in a commercial industrial context, requires the exploration of these related topics.

Figure I-I is a macro level mind-map (Buzan and Buzan, 1993) developed initially in the planning stages of this work and is therefore derived from the perspective of the empirical study (Part II). Although this is a very general representation, it illustrates the topics associated with training evaluation and their relationship to it.

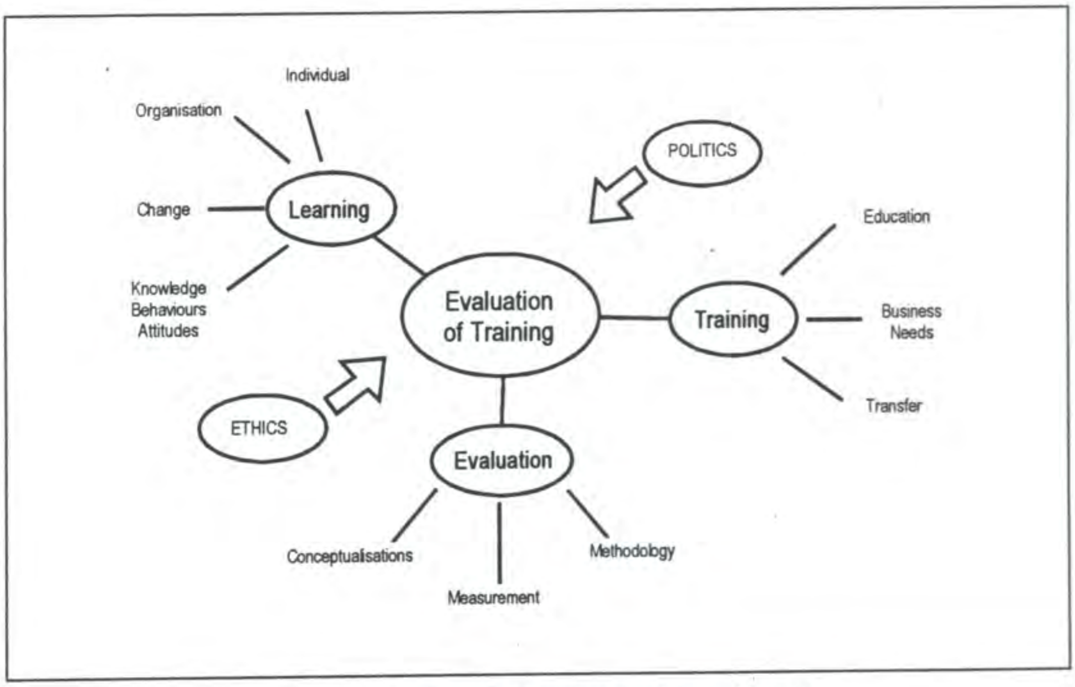

Figure I-I: Topics associated with the evaluation of training

In simple terms, the evaluation of training relates to the bodies of knowledge concerned with; learning; training; and evaluation. Within each of these bodies of knowledge, key concepts can be distinguished (i.e. learning can be distinguished with respect to; the individual person; the organisation; change; and knowledge, behaviours, and attitudes).

In addition to the related bodies of knowledge, the evaluation of training is influenced by a range of social factors and for the purposes of this thesis, politics and ethics are identified as being significant to the practice of evaluation of training.

The purpose of this review of the literature is to explore these general topics, giving particular attention to those areas which more directly relate to the focus of this study; the conceptualisation of training evaluation. This part of the thesis comprises of three chapters; chapter 1 - learning and training; chapter 2 - conceptualisations of evaluation; and chapter 3 - measurement and methodology.

Learning and Training

Learning and Training are themselves inextricably linked and are pertinent to the evaluation of training. The goal of training is to direct, facilitate and accelerate learning in support of organisational objectives. To comprehend training requires an understanding of (at least) the basic principles of learning.

Conceptualisations of Evaluation

To understand the conceptualisations of evaluation, it is necessary to consider the wider concept of scientific inquiry and how it relates to evaluation. In scientific enquiry, whether set in a positivist or post-positivist paradigm (these paradigms are discussed later), the goal of the inquirer is to understand phenomenon.

Within these paradigms of inquiry exist conceptualisations of how evaluation is undertaken. A conceptualisation is a general notion; a theme or a design, of an activity (Brown, 1993). Conceptualisations of evaluation have evolved over centuries. Planned evaluation is documented as early as 2200BC with the selection of personnel in China (Shadish et al, 1995), however, this review is primarily concerned with the conceptualisations of evaluation since the 1940's, starting with Tyler, acclaimed to be the 'father of evaluation' by the Joint Committee on Standards for Educational Evaluation (1981).

Political Factors

The many parties involved in any social programme, from training in a commercial context to good cause awareness in a voluntary sense, often have interests in the role, importance, and continuation of the programme. Rossi and Freeman (1993) describe the evaluator's role as an expert witness testifying to the degree of a programme's effectiveness within a political system that is sensitive to weighing, assessing, and balancing the conflicting claims and interests of a number of constituencies.

Politics is a valuing activity, with the value criteria varying between interested parties in support of their own overt or covert aims and intentions. Evaluation therefore engages in the political process involving multiple stakeholders and contributes knowledge to the decision making process. Because politics relates to and influences the overall conceptualisation of evaluation, the topic is reviewed as part of chapter 2.

Measurement and Methodology

Across the range of conceptualisations of evaluation presented in the literature, measurement provides a universal way of describing, comparing and valuing phenomena. Methodology and the techniques of evaluation vary across conceptualisations, but they provide the means by which data is collected and analysed. The purpose of this chapter is to provide an overview of measurement and evaluation methodologies, paying particular attention to those methods employed in the empirical study.

Ethical Considerations

The purpose of data collection and its subsequent analysis is to gain a true understanding of phenomena, that is to say one which accurately reflects reality. This relies on the accuracy of the data collected which is dependent on the honesty of responses given, particularly in the areas of interviews, questionnaires, and observation.

If accuracy of information was the singular objective of an evaluative investigator, then by collecting data without subjects (respondents) knowing or consenting would increase the likelihood of accurate data collection. Douglas (1976) argues that the only way to gather data of any validity is for the researcher to operate in a covert manner. For researchers, including those involved in evaluation, however, ethics are an important consideration as they have moral, legal and professional implications.

Ethical issues relate to many aspects, including accuracy, confidentiality, breadth of consultation, rights of consent and access, and continuity of purpose (Raffe et al, 1989). Morrison (1993) argues that the right of the public or stakeholders to have access to evaluation data and the right of the individual to privacy is a fundamental tension which is central to the ethical issues surrounding evaluation.

Chapter 1

Learning and Training

1. Learning & Training

> *'Experience is the child of Thought, and Thought is the child of Action. We cannot learn men from books alone.*
>
> *Disraeli, 1826*

1.1 Introduction

This chapter is divided into two parts; learning; and training. The broad spectrum of learning is reviewed and particular emphasis is given to those areas which specifically relate to this study. An overview of training and education is undertaken which, together with learning, provides the training context for the review of the evaluation literature.

Despite the vast array of literature with respect to learning and training, the engineers' quality improvement training programme was not developed from any theoretical base. Whilst this may appear to be unusual, many writers have commented that the theory and practice of training are not well integrated. In the last thirty years this non-integration has been a recurring theme in the literature. Campbell (1971) commented that whilst training and development literature was voluminous, it was largely non-empirical. Wexley (1984) offered a similar conclusion, noting that large areas of training, in particular the factors which facilitate the transfer of training to the workplace, were in need of empirical study. Latham (1988) concluded that the training literature had become more theoretical, but that training practitioners largely ignored the results from research literature. Cannon-Bowers and Tannenbaum (1991) observed that theories of learning and training are not generally integrated with the practice of training, and that research findings are not translated into useful training methods.

As part of this review, relevant aspects of the literature are retrospectively related to the engineers' training programme in an attempt to provide its theoretical grounding.

1.2 Learning

The modern literature of learning is vast; spanning the behavioural-associationist theories of; Thorndike (1913); Pavlov (1927); Skinner (1938); and others, to the cognitive-organisational theories of; Koffka (1924); Tolman (1932); Feigenbaum and Feldman (1963); and others, to the neurophysiological theories of; Pitts and McCulloch (1947); Grossberg (1978); and others. Kim (1993) observes that learning has been an important subject for research for psychologists, linguists, educators and others, yet despite all the research, relatively little is known about the human mind and the learning process. Whilst learning is a multifaceted concept and an important aspect with regard to the evaluation of training, for the purposes of this study the review is limited to those areas which have been directly related to adult training and work organisations in the literature.

1.2.1 A Definition of Learning

A definition is, according to the Oxford English Dictionary (Brown, 1993), "a precise statement of the nature, properties, scope, or essential qualities of a thing; an explanation of a concept etc.; a statement or formal explanation of the meaning of a phrase". These statements presuppose the general agreement on what something is. In the case of learning, the differing theoretical explanations make it impossible to give 'a precise statement of the nature, properties, scope or essential qualities' of learning which satisfy the interpretations of the various schools in the literature. For the purposes of this study, an operational definition of learning from Bower and Hilgard (1981: p.11) is adopted:

> *"Learning refers to the change in a subject's behaviour or behaviour potential to a given situation brought about by the subject's repeated experiences in that situation, provided that the behaviour change cannot be explained on the basis of the subject's native response tendencies, maturation, or temporary states (such as fatigue, drunkenness, drives, and so on)."*

Whilst the definition is written in the wider sense of all learning, not just that restricted to humans, it is applicable to this study. The definition is centred on the learner and the important features of Bower and Hilgard's definition are:

(i) The notion of change in behaviour as learning is a hypothetical construct; that is to say that it cannot be directly observed but can only be inferred from observable behaviour (Gross, 1992).

(ii) The explicit exclusion of the subject's behaviours due to 'native response tendencies' from learning. Native response tendencies refer a person's natural instincts (i.e. the instinct to eat when hungry, or to shiver when cold).

(iii) The notion of non-temporary states of behaviour change. Permanency as a feature of learning is a recurring theme among many writers; Atkinson et al (1993) - "Learning may be defined as a *relatively permanent change* in behaviour that results from practice."; Coon (1983) - "A *relatively permanent change* in behaviour due to past experience."; and Kimble (1961) - "A *relatively permanent change* in behavioural potential which accompanies experience but which is not the result of simple growth factors or of reversible influences such as fatigue or hunger."

This third feature has particular significance to Kirkpatrick's conceptualisation of evaluation (Kirkpatrick, 1959, 1960, 1994), which is reviewed in chapter 2. Kirkpatrick's interpretation of learning does not feature permanency, and given the influence of Kirkpatrick on training in the last 30 years, this has wide ranging consequences for how training and its relationship to learning are perceived.

1.2.2 Theories of Learning

Kerlinger (1986) defines theory as 'a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purposes of explaining and predicting phenomena'. A theory therefore is an explanation of phenomena in terms of its variables. Theory is important to science and research as it provides general explanations and the establishment of general laws; Braithwaite (1955) defines the purpose of science as 'establishing general laws covering the behaviours of the empirical events or objects with which the science in question is concerned, and thereby to enable us to connect together with our knowledge of the separate known events, and to make reliable predictions of events as yet unknown'.

Patrick (1992), from an extensive review of the literature, identifies five theories of learning which are pertinent to training; (i) three-phase theory (Fitts, 1962); (ii) Anderson's theory of cognitive skills acquisition (1982, 1983); (iii) MacKay's theory (1982); (iv) closed loop theory of motor learning (Adams, 1971); and (v) schema theory of motor learning (Schmidt, 1975).

Of these five theories identified by Patrick, Fitts' three-phase theory and Anderson's theory of cognitive skills acquisition have particular relevance to this study. Adams' closed loop theory of motor learning and Schmidt's schema theory of motor learning fall outside the scope of this thesis as the evaluand considered is not concerned with motor-skills learning and training. MacKay's theory, which was developed in the area of speech production, is also not included as this has not been generalised to other areas of learning and training in the literature.

It should be noted by the reader that in the context of this study, skill is a wide ranging concept and includes cognitive, perceptual, motor, and social skills. Furthermore, knowledge is an antecedent of skill. More contemporary training literature (Walker, 1992; Reid and Barrington, 1994; Senge et al, 1994) refer to behaviours and competencies, as opposed to skill. Within the context of this thesis, these fall within the reference of skill and so skill is defined in the sense of Argyle (1967) and Jackson (1989):

> *Skill is the capability to perform a range of functions with ease and precision through; knowledge of what is required; the ability to translate knowledge into specific behaviour; and a motivation to perform.*

1.2.2.1 Fitts' three-phase theory

Fitts (1962) postulates that the development of skill progresses through three overlapping stages of; cognition; fixation / association; and autonomous. The cognitive phase is concerned with the acquisition of knowledge and skills to a level where the learner is able to verbalise what (s)he has learned. The fixation or associative phase is concerned with practising the skills acquired in the first phase and the making and eliminating of errors. This phase often overlaps with the preceding phase. The final phase is that of autonomy where, through the learner's repeated practice of the skills until they become automatic, skills performance requires fewer of the learner's psychological resources such as memory and concentration. Once again, Fitts proposes that this phase overlaps with the preceding phase.

1.2.2.2 Anderson's theory of cognitive skill acquisition

Anderson (1982, 1983) proposes three main stages in the learning of a skill; the declarative stage; the knowledge compilation stage; and the tuning stage. The declarative stage is concerned with the learner *knowing facts* relevant to the desired skill. Skill performance is attempted by the learner using these facts. By way of illustration, Anderson cites the example of learning to drive a car where the facts are knowing where the gear lever is and what is does. Anderson postulates that as skill performance is dependent on the learners working memory, performance will initially be slow, inaccurate and effortful.

The knowledge compilation stage is concerned with the development, by the learner, of specific procedural knowledge. Procedural knowledge is *knowing how* to perform a skill. This learner translation of declarative knowledge into procedural knowledge through practice or experience is termed by Anderson as knowledge compilation. Knowledge compilation is the learner's development of rules which govern procedures. The processes of knowledge compilation are; *composition* where adjacent rules are collapsed or merged into more direct rules; and *proceduralisation* where a rule incorporates more task relevant information.

The tuning stage is where the learner further improves procedural knowledge by tuning or adjusting skill performance to suit a range of situations. This will involve generalising rules to suit a wider range of applicability, discriminating between rules which have been successfully and unsuccessfully applied in different situations, and strengthening rules which are used and weakening rules which are not used or used unsuccessfully.

Whilst similar to Fitts' three-phase theory, Anderson's theory has elaborated Fitts' ideas in greater detail and has provided a framework for understanding a range of skills including computer programming, text editing making geometry proofs, performing arithmetic and playing chess (Patrick, 1992). Hunt (1989) observed that the range of applications of the theory is impressive.

With respect to the empirical study, the design of training for all content areas of the programme followed a general model of; (i) presentation (lecture); (ii) content item practice exercise; (iii) module content practice exercise; and (iv) application to the engineering process. Contrasting this basic model with Anderson's theory of skill acquisition, the first element (i) of the model relates to Anderson's declarative stage of skill acquisition and, to some extent, to the compilation stage. The purpose of the presentation is to provide trainees with knowledge of the given topics by defining principles and the framework of the methodology. These are set in the context of engineering using simple examples or case studies. In this sense, the presentation provides a procedural framework to the learners for the development of the compilation stage of their learning.

Each content item is reinforced in the training through the use of a learner exercise. The exercises, normally defined in terms of a scenario or problem, provide the context for, or input to, the use of the content item (e.g. customer product requirement information and alternative design concepts for a corkscrew are provided and learners are asked to select the best design concept using Pugh concept selection - see chapter 4). In terms of Anderson's theory, this is intended to facilitate the compilation stage of learning.

The theories identified by Patrick are concerned with the internal mechanisms of the learner and conceptualise learning as a progressive process. For the purpose of this study, the inference to be drawn is that learning in the initial stages can be directed and stimulated by training, however it should be acknowledged that training is not the only factor in this process, and that training in the sense of the engineers' quality improvement training programme cannot directly affect the final stages of learning. For both Fitts and Anderson, this final stage of the process is dependent on 'the learner's repeated practice of the skills until they become automatic' and 'adjusting skill performance to suit a range of situations' respectively. Any assessment of training effectiveness with respect to learning can only be done after trainees have had the opportunity to fulfil these criteria.

Another important aspect of learning, with respect to training, is that of the influence of the environment. Bandura's (1977) social learning theory looks outside of the individual at information exchanges with others to explain learning. Latham (1989) acknowledges Bandura's social learning theory as having an immediate and positive influence on training and development programmes.

1.2.2.3 Social Learning Theory

The central idea of social learning theory is that an individual learns from another by means of observational modelling. Observational modelling is where a person observes what another person is doing and then does something similar. The theory states that behaviour is a continuous reciprocal interaction among cognitive, behavioural and environmental factors and therefore behaviour change is both determined by and affects environmental consequences. Two individual difference factors were found (Bandura et al, 1977) to affect behaviour change, with respect to social learning theory. These were self-efficacy (a person's belief that (s)he can perform a given behaviour in a given setting) and outcome expectancies (a person's belief that the given outcome will occur if (s)he engages in the behaviour).

The emphasis is on observation and imitation in acquiring new behaviours and in regulating the frequency and occasions of their appearance. With respect to training, this notion is significant to the reinforcement of learning back in the workplace. Where a trainee returns from a training course to an environment which displays the behaviours taught in the course, further learning is likely to occur as the trainee observes and imitates the behaviours of colleagues (Latham, 1989). If however, the trainee returns to an environment where the taught behaviours are not observed, or counter behaviours are evident, this will affect the trainee's willingness to practice the behaviours taught on the course and exclude the opportunity for further learning through observation and imitation.

If Bandura's social learning theory widens the perspective on learning and its significance to training, then the evolving concept of the learning organisation further extends this perspective.

1.2.3 Learning Organisation

Many writers have conceptualised learning in an organisational sense. Popularised by Peter Senge (1990), the term 'learning organisation' has proliferated much of contemporary training and development literature. The foundations for the notion of a learning organisation were laid by Argyris and Schon (1978), who developed the concept of single- and double-loop learning in an organisational context. They regarded learning as involving the detection of errors and their subsequent correction. Where detection and correction enabled the policies and objectives of an organisation to continue, they labelled the process single-loop learning. If, however, the detection and correction activities changed policies and objectives, Argyris and Schon described it as double-loop learning as it involved learning from others and a willingness to accept change. Sugarman (1998), referring to this distinction first order (single-loop) and second order (double-loop) learning, comments that first order learning occurs within a framework of customary, accepted assumptions, while second order learning questions those assumptions.

Peddler et al (1991) define a learning organisation as one which "facilitates the learning of all its members and continually transforms itself". Honey & Mumford (1996) elaborate on this definition, describing a learning organisation as "having managers who create an environment where the behaviours and practices involved in continuous development are actively encouraged". Honey and Mumford go on to identify ten behaviours that increase learning; asking questions; suggesting ideas; exploring options; taking risks / experimenting; being open about the way it is; admitting inadequacies and mistakes; converting mistakes into learning; reflecting and reviewing; discussing what has been learned; taking responsibility for own learning and development.

Learning organisation, therefore, is a term which describes acceptable behaviours within an organisation which are conducive to learning. It places an emphasis on the responsibility of individuals for their own learning and a general acceptance of this responsibility not only encourages many individuals to learn, but has an effect on the learning of others. This, in turn, affects the individual's learning. Hammond and Wille (1991) describe this concept of learning organisation as "synergystic" in that the learning by the organisation is greater than the sum of the learning of its individuals.

1.2.3.1 The Learning Cycle

Sloman (1994) identifies the work of Kolb (1984) as making a significant contribution to the development of the learning organisation. Kolb introduced the concept of the learning cycle where learning occurs through a 4 stage cycle of; 1) having an experience; 2) observing and reflecting on that experience; 3) developing principles and concepts from the reflection; and 4) testing the principles and concepts by either replicating the original experience or trying them out in a new similar experience, which becomes the experience in stage 1. Proponents of the learning organisation (e.g. Sloman, 1994; Honey & Mumford, 1996) argue that the individual's experience of the learning cycle can also be paralleled in the organisation.

Therefore the significance of the concept of a learning organisation on individual learning with respect to training are two fold:

(i) Where training is in support of organisational goals, learning is likely to be reinforced and further learning encouraged after completion of training. Kandola (1993) acknowledges the influence of others on an individual's learning, arguing that an environment and culture which is tolerant of risk-taking and error-making encourages individuals to attempt new skills. Barham et al (1988) describe the learning organisation as one which is not restricted to discrete training events, but one where it has become a continuous process and on-the-job learning is a way of life.

(ii) Trainees bring learning skills and a sense of their own responsibility for learning to training.

Calvert et al (1994) found from an extensive survey of HRD professionals and line managers from USA companies that the concept of a learning organisation forces a training department to be unusually flexible and expert in offering just-in-time training.

Jones and Hendry (1994) argue that by widening our understanding of the contexts in which learning occurs we begin to provide a framework for greater organizational and individual learning capability. Jones and Hendry observe that learning is as much acquired through access to a range of activities, ideas and skills that broaden a learner's understanding of a variety personal, social and employment networks, as it is through formal training. These additional characteristics of learning are referred to by Jones and Hendry as 'soft' learning characteristics, whereas 'hard' learning is pragmatic, formal and brought about through prescribed training. Soft learning is often unintended, indirect, and not controlled by the organisation, yet it is central to Knowles' (1990) contention that adults prefer self directed learning, learn most effectively through experience, and by means of actual day to day jobs and routines.

The significance of 'soft' learning to training is made clearer by Jones and Hendry in their argument that soft learning is concerned with the social contexts in which (trained) technique is applied and developed. This relationship between 'soft' learning and the transfer of training to the workplace has not been made explicit in the literature relating to transfer of training (section 1.3.2).

The basis for Jones and Hendry's argument is largely attributed to the notion of adult learning as conceived by Knowles (1990). Knowles contends that adult learning is characterised by personal autonomy, experiential learning, people seeing connections between different aspects of their work and lives, and by activities that may initially seem irrelevant to the specific job being undertaken. Adult learning has been the subject of many writers (e.g. Dubin and Okun, 1973; Simpson, 1980; and Brookfield, 1986) who all place emphasis on maturity, past experiences and a tendency toward self directedness in their learning as distinguishing factors from younger learners.

The significance of the learning organisation to training and the evaluation of training can be summarised in terms of the learners' context. Where a training programme is undertaken for learners who come from an environment which is characterised as a learning organisation, the inference is that they are more likely to take risks and attempt to apply new skills and co-workers are more likely to observe and learn from trainees. In this sense, the organisation is a critical factor to both the learning process and therefore the effectiveness of the training programme.

Factors of learning, whether environmental in the sense of Bandura and Senge et al or internal to the learner, are considered by Gagné (1970) in terms of conditions of learning.

1.2.4 Conditions of Learning

Gagné (1970) classifies human learning into five categories or domains; verbal information; attitudes; intellectual skills; motor skills; and cognitive strategies. The basis for Gagné's classification is that learning occurs when an individual acquires a capability to do something, and as the learned capability is not directly observable, it is from the learner's behaviour that learning capability is inferred. The behaviours are outcomes of learning and different learned capabilities result in correspondingly different outcomes, hence the basis for his classification.

For each category or domain of learning, Gagné proposes that a whole set of factors influence learning, and collectively described as internal and external conditions of learning. Internal conditions of learning refer to the acquisition and mental storage of prior capabilities that are either essential to, or supportive of, subsequent learning. Cormier and Hagman (1987) describe these conditions as transfer of learning or transfer of training. These are sometimes referred to in the literature as the learner's states of mind (e.g. Patrick, 1992). External conditions refer to the various learning stimuli that are in the learner's environment. These convey the (new) information; such as knowledge and skills.

From a training perspective, conditions of learning and the relationship between the internal conditions and the external conditions are significant. External conditions can be manipulated by the course design to activate and support the internal processes of learning. By recognising and understanding that different sets of conditions are needed for learning to occur for the different domains, appropriate training designs can be conceived for the appropriate domain. Conditions of learning, therefore have great significance for training. Aronson and Briggs (1983) acknowledge Gagné's work, along with that of Briggs (1970, 1977), as greatly contributing to what is known about human learning that is relevant for instruction (training).

From this general review of the literature thus far, the inference to be drawn is that learning is conceptualised as an abstract concept which can be inferred by changes in knowledge and skill. Changes can be brought about by the learner's exposure to external conditions such as observation of others, statements of information and the opportunity to practice. These conditions alone, however, are not sufficient for learning to occur. Learning is also affected by the learner's internal conditions, in both the sense of Gagné et al and the learner's willingness to learn. Motivation and self efficacy are important factors to learning and, therefore, to training and its evaluation.

1.2.5 Aspects of the Learner

The mechanistic theories of learning proposed by Fitts and Anderson and subsequently widely adopted in the design of training pay little attention to internal aspects of the learner. Gagné, to some extent, takes account of the learner's characteristics with respect to learning, however these are not elaborated to any real extent. Bandura refers to self-efficacy; the learner's belief in his/her capability to learn and perform a task.

Noe and Schmitt (1986) contend that learner characteristics which influence learning are of utmost importance. Wexley and Latham (1981) define learners' characteristics in terms of their trainability which they define as a function of the trainee's ability and motivation, however Noe (1986) citing Porter and Lawler (1968) extended the definition of trainability as a function of ability, motivation and perceptions of the work environment. Maier (1973) identified motivation as a critical factor to learning in a training context, indicating that performance in training will be poor if motivation is low or absent, however Wexley (1984) observed that the majority of trainability studies have focused on ability factors, or as he terms 'can do' factors, as opposed to motivational factors, or 'will do' factors.

More recently Honey and Mumford (1986 and 1992) defined four major categories of learning styles which distinguished preferences of different learners. From the literature, therefore, four aspects of the learner can be identified as being significant to learning, and hence to training and its evaluation. These are; motivation; ability; perceptions of the work environment; and preferred learning styles.

1.2.5.1 Motivation.

Steers and Porter (1975) suggested that motivation is composed of energising, directing, and maintenance components. In a learning situation, Steers and Porter argue that motivation is the force that influences enthusiasm about the programme (energizer); a stimulus that directs participants to learn and to attempt to master the content of the programme (director); and a force that influences the use of newly acquired knowledge and skills (maintenance).

With regard to the effect of the work environment on motivation to learn and apply knowledge and skills to the job, O'Connor et al (1984) suggests that trainees' perceptions of task constraints, such as lack of equipment or financial resources, may indirectly influence behaviour change and learning by either reducing motivation to learn new skills or application of skills acquired in training to job tasks.

1.2.5.2 Ability

Ability is defined an terms of a learner's capacities related to the performance of some set of tasks (Fleishman and Mumford, 1989). These general capacities vary in individuals and can develop over time through the interplay between genetic influences and the cumulative effects of prior developmental experiences (Lohman and Snow, 1984).

The significance of learner ability on training is well documented. Noe (1986) concluded from a review of the literature that most studies addressing trainability had focused on learner's ability. Patrick (1992) distinguishes abilities in terms of cognitive and motor abilities and observes that many different ability taxonomies exist within the same areas concluding that whilst none are incorrect, each is more useful in a different context.

Fleishman and Mumford (1989) comment similarly and from an extensive review of the literature identify fifty learner ability constructs for consideration in learning and training. These are too numerous to list here, but include, for example, constructs such as oral comprehension, memorizatrion, and perceptual speed. Fleishman and Mumford conclude that individual capacities that trainees bring are a critical determinant to the success of training efforts.

1.2.5.3 Perceptions of the Work Environment

Clark et al (1993) found that learners in an organisational context who did not believe their supervisors would support the application of new learning to the job had a negative effect on their willingness to learn. Reid and Barrington (1994) and Sloman (1994) draw similar conclusions from their extensive reviews of the literature and contend that the management of work-environment conditions with respect to training is an area much neglected by most training departments. Reid and Barrington's analysis extends beyond supervision to include all workers within the organisation and draw parallels to the concept of learning organisation.

1.2.5.4 Learning Styles

Building on the work of Kolb (1974), Honey and Mumford (1986, 1992) distinguished learner preferences into four categories; (i) activists; (ii) reflectors; (iii) theorists; and (iv) pragmatists. Honey and Mumford derive their categories with respect to experiential learning and in the context of Kolb's (1974) learning cycle.

Activists are characterised by their desire to fully involve themselves in new experiences and tending to have few biases. Honey and Mumford describe them as 'enjoying the here and now and happy to be dominated by experiences'. Other characteristics of activists include their open-mindedness and enjoyment of challenges. They tend not to thrive on longer term implementation and consolidation of experiences.

Reflectors are characterised by preferences to collect and analyse information from a range of experiences and events before reaching definitive conclusions. They have a preferences for viewing problems or experiences from different perspectives and tend to be cautious and thoughtful before committing to an idea. Reflectors characteristically adopt a low profile, preferring to listen to others as opposed to taking the lead in discussions.

Theorists are characterised by assimilating disparate facts into coherent theories. They think problems through in a vertical, step by step logical way. They prefer basic principles and models and their thinking rationale is based on logic. Subjective judgements and lateral thinking are areas which theorists tend to find uncomfortable.

Pragmatists are distinguished by their preference to apply ideas which attract them to see if they work in practice. They regard problems as challenges and look for the practicality of new concepts and ideas. Pragmatists tend not to value open ended discussions or philosophical debate which do not yield actionable outcomes.

Whilst Honey and Mumford derive their categories with respect to experiential learning, Reid and Barrington (1994) observed that the notion of learning styles is an important factor to the design of training. Different learners respond better to different instructional methods and therefore, as part of a training needs analysis, the types of learners should be identified and their preferences reflected in the design of training.

As trainers, we have no direct control or influence over the outcomes of training; we can only control and influence the training process and the inputs to the process. With regard to the evaluation of training, measuring outputs in the sense of Kirkpatrick provides little information about the training process or the inputs to the training. With a given, or limited, level of resource, information on inputs and process is far more valuable in terms of the management and improvement of the programme than that on its outputs (see chapter 7 for analysis and discussion).

1.2.6 Relationship of Learning to Training

Gagné, Briggs and Wager (1992) argue that training should stimulate individuals in such a way as to bring about desired changes in behaviour. The process that makes such change happen is called *learning* and the situation that sets the process into effect is called a *learning situation*. The relationships between learning situations and behaviour change are referred to by Gagné (1970, 1985) as the *conditions of learning*. These are the conditions, both internal and external to the learner, that make learning occur.

The inference to be drawn from the literature with respect to training in the context of this study is that whilst learning is the principle aim of training and, as such, provides the basis for evaluating training effectiveness, it is not exclusive to training. Learning occurs within and outside of training. Individuals learn from others in the sense of Bandura, and as part of an organisation where there exists an environment and culture which is encouraging to learning by being tolerant of risk-taking and error-making as in Peddler et al (1991). A training programme can shape the learning of an individual within an organisation without that individual participating directly in the training programme.

1.3 Training

From the general introduction, the reader will have recognised that context is a critical defining factor of evaluation. The general context for the subject of this research are the fields of education and training. The specific topic of the evaluation is the training programme itself, which is considered in detail in Chapter 4.

The engineers' quality improvement training programme is described colloquially as a training programme. This is based on the popular perception within the company that the Education & Training department provide a training service, with the exception of sponsored undergraduate and post graduate degree programmes which are considered to be education. The following sections review the literature in an attempt to understand and distinguish education from training.

1.3.1 Definitions of Education and Training

Since joining the Education and Training department in 1987, and perhaps even before, I have pondered the question "what is 'education' and what is 'training' and what is the difference?" This question has never been answered satisfactorily. My first training manager told me that 'education' was concerned with generality and 'training' was specific - a view shared by Glaser (1962) who used the specificity of objectives to distinguish training from education. This in itself does not provide a clear distinction; at what point does generality become specificity?

Definitions of education and of training commonly referred to in the literature (e.g. Patrick, 1992; Bramley, 1991) are those of the Department of Employment (1971). The Department of Employment define education as:

> *'Activities which aim at developing the knowledge, moral values and understanding required in all walks of life rather than knowledge and skills related to only a limited field of activity. The purpose of education is to provide the conditions essential for young persons and adults to develop an understanding of the traditions and ideas influencing the society in which we live, of their own and other cultures and of the laws of nature, and to acquire linguistic and other skills which are the basic to learning, personal development, creativity and communication.'*

This definition draws a distinction between 'knowledge, moral values and understanding required in all walks of life' and those 'related to a limited field of activity'. The implication is, similarly to Glaser, that generality distinguishes education from training. This is reinforced by the latter part of the definition which describes a wide ranging purpose to education. This description of the purpose of education relates the definition to the wide ranging curricula of education, and not just singular activities of education.

With regard to training, the Department of Employment (1971) offer the following definition:

> *'The systematic development of the attitude / knowledge / skill behaviour pattern required by an individual in order to perform adequately to a given task or job'.*

By this definition, training is conceptualised as being 'systematically developed' implying planning and controlled activities to develop an individual's attitudes, knowledge and skills within a specific task or job context. It focuses on the individual (whereas education is defined in terms of groups) and on a given job or task, where performance is the criterion for success.

The inference to be drawn from these definitions is that training is centred on job or task performance of individuals, whereas education is concerned with the whole arena of knowledge in all walks of life for all people.

French (1989) draws similar conclusions from the literature. Citing definitions given by Moore (1982), French observes; (i) that education is distinguished from training by its generality; and (ii) that a particular approach to learning is distinguished between education and training. French argues that education is held to an organic view (one of personal growth) and training to a mechanical view (one of behavioural expression).

More recently however, Reid and Barrington (1994) have observed that the terms 'education' and 'training' are both frequently used in training literature and suggest 'that perhaps the nation is beginning to bring together its education and its training activities' (p.42). Training has evolved significantly in recent years in order to keep pace with technological advances and the increasing levels of skills demanded of the workforce (Ashton and Green, 1996). The traditional narrow view of training no longer holds for many areas of the workforce.

Returning to the Department of Employment's definitions from this perspective, the definition of training is more specifically framed than for that of education. It does not refer to the wide arena of all training for all people, as is the case for education, but for a specific task or job for an individual. In this sense, it is the frame of definition, as opposed to the concept it conveys, which is more specific for training and more general for education.

It is my contention that the distinguishing factors of education and training are diminishing as training continually develops its learning strategies and expands its curricula. The similarities between education and training outweigh the differences and therefore, for the purposes of this study, the review of the literature with regard to evaluation in chapter 2 deliberately crosses the boundaries into evaluation of education with the intention of providing a more comprehensive and informative review.

One significant difference, however, between education and training in a general sense, is that of the relationship between the traditional view of training and the performance of work. Learning is directed at bringing about changes in job performance, and this relationship is the focus of transfer of training.

1.3.2 Transfer of Training

The term 'transfer of training' is widely used within education and training literature. The term is used to reflect a wide range of concepts including; acquisition; performance; and relearning of knowledge, skills and attitudes during and following training.

Buckley and Caple (1992)) define transfer of training:

> *'Transfer of training occurs whenever the existence of a previously established habit or skill has an influence on the acquisition, performance or relearning of another habit or skill. In the training context positive transfer will have taken place if the trainee is able to apply on the job what has been learned in training with relative ease or is able to learn a new task more quickly as a result of earlier training on another task. Conversely negative transfer arises when performance on the job or on the new task is decelerated or hindered by what knowledge and skills have been acquired.'*

Buckley and Caple characterise transfer of training in two respects; (i) as the influence of previous learning on new learning; and (ii) as the application of trained skills to the job. Transfer of training, therefore takes place both during the learning of new skills and in the performance of those skills.

The notion of previous learning influencing new learning, is also termed 'transfer of learning' by Cormier and Hagman (1987), who use the terms 'transfer of learning' and 'transfer of training' interchangeably.

The literature can be distinguished as centring on two main topics; transfer with respect to acquiring (learning) new skills; and transfer with respect to applying new skills. The former features largely in the literature of learning, which was reviewed previously. The latter is the focus of many training writers and has particular importance with respect to the evaluation of training.

The application of trained skills forms the focus of definitions offered by several writers. Broad and Newstrom (1992) define transfer of training as "the effective and continuing application, by trainees to their job, of the knowledge and skills gained in training"; Wexley and Latham (1981) define transfer of training as "the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job"; Baldwin and Ford (1988) offer a similar definition as "for transfer to have occurred, learned behaviour must be generalised to the job context and maintained over a period of time on the job"; and Phillips (1991), "transfer of training refers to the extent to which the learned behaviour from the HRD [Human Resources Development] program is used on the job".

Broad and Newstrom (op.cit.) and Baldwin and Ford (op.cit.) make explicit another dimension of transfer of training; that of continuing application. Patrick (1992) terms this as 'retention' and goes on to identify factors affecting retention, commenting on the high level of agreement within the literature.

1.3.2.1 Factors affecting Transfer and Retention.

Patrick concludes from the literature that retention of skills is positively related to the level of learning at the end of a training course. Citing Gardin and Sitterly (1972), Patrick identifies three generalisations from the literature; (i) performance level at the end of training; (ii) duration of retention interval (between training and using the skills); and (iii) skills rehearsal. Generalisations (ii) and (iii) are both concerned with timeliness of training given to trainees.

Patrick found literature to support the notion that where a trainee achieved a high level of skill performance at the end of training they were more likely to retain the skills for a longer period of time than those who didn't master skills performance. With regard to retention intervals and rehearsal, not surprisingly, Patrick found evidence to suggest that the longer skills were not used (retention period) the greater the likelihood they would be lost by the trainee. With respect to training delivery and to training evaluation, timeliness of training is an important factor to training outcomes.

Baldwin and Ford (1988) identified factors of transfer in terms of training inputs and environmental characteristics from their extensive review of the organisational training literature. They concluded that trainee characteristics such as ability, personality and motivation; and training design in terms of transfer strategies employed are significant factors to the generalisation and maintenance of trained skills on the job.

1.3.2.2 Transfer Strategies

Broad and Newstrom (1992) describe the aim of training as being to maximise the amount of positive transfer. Baldwin and Ford (1988) identify four basic strategies for the transfer of training from an extensive review of empirical research; identical elements; general principles; stimulus variability; and conditions of practice.

The notion of *identical elements* was originally proposed by Thorndike and Woodworth (1901) who hypothesised that transfer is maximised to the degree that there are identical stimulus and response elements in the training and transfer settings. Baldwin and Ford found empirical research which supported the use of identical elements as a strategy of increasing the retention of motor skills and verbal skills.

Training through *general principles* maintains that transfer is facilitated when trainees are taught the general rules and theoretical principles that underlie the training content which is applicable to the job. Again, Baldwin and Ford found empirical evidence to support this strategy, particularly with respect to cognitive skills (e.g. Crannell, 1956; and Goldbeck et al, 1957).

Stimulus variability is the notion that positive transfer is maximised when a variety of relevant training stimuli are employed. The use of several examples of a concept strengthens trainees' understanding so that they are more likely to see the applicability of a concept in a new situation. In additional to the empirical support for the strategy of stimulus variability found by Baldwin and Ford, Cormier and Hagman (1987) identify several studies (e.g. Gick and Holyoak, 1983; and Homa and Cultice, 1984) which support the hypothesis that positive transfer increases with stimulus variability.

Conditions of practice subdivide into several strategies for training design; massed/distributed training; whole/part training; feedback; and over-learning. *Massed/distributed* training concerns dividing training into segments, with Baldwin and Ford observing research evidence to suggest that distributed training is more likely to be retained longer. *Whole/part* training concerns the practising by trainees of content material either as a whole, or as constituent parts. From the research, Baldwin and Ford conclude that the practice of the whole is more effective with regard to transfer when either trainee intelligence is high; training is distributed, rather than massed; or the training material has a low (simple) level of complexity.

Feedback refers to information provided to trainees about their performance both within training and after training. Baldwin and Ford found evidence to support feedback as being critical. *Over-learning* refers to the process of providing trainees with continued practice far beyond the point when the skill has been performed successfully. The hypothesis is that the greater the amount of over-learning, the greater the subsequent retention of the trained material. Again Baldwin and Ford found considerable evidence in support of this.

1.3.2.3 Transfer Problems

Transfer of training , or the lack of it, is the attention of many writers. Hoffman (1983) reported estimates that only 10% of expenditures for training resulted in observable behaviour change on the job. Similarly, Baldwin and Ford (1988) concluded that there was a growing recognition of a 'transfer problem' and estimated that USA industries spend up to $100 billion dollars annually on training and development with less that 10% of expenditures resulting in transfer to the job. However, Phillips (1991), who estimates the total expenditure on training in the US to be at approximately $100 billion, observes that the lack of evaluation directed at measuring 'bottom line results' hinder's measurement of transfer in financial terms.

Broad and Newstrum (1992) identify transfer of training as a problem, but conclude that there is relatively little empirical research on transfer problems. With respect to evaluation of training, transfer problems have particular significance. Much contemporary literature emphasises the need for evaluations to measure training in terms of its contribution to the business. Several writers (e.g. Phillips, 1994; and Basarab and Root, 1994) have described this in terms of a *return on training investment* for organisations.

With regard to the evaluation of training, transfer of training (as the application of learned skills to the job) occurs away from the training environment; after the trainee has finished with training and returned to the job. Transfer of training is a problem, yet it is unclear from the literature as to whom the problem belongs. Training is the concern of the training department, yet this is often perceived as merely providing training courses which potentially meet the organisations needs (Brinkerhoff, 1991).

The inference to be drawn from the literature is the criticality of transfer of training to the achievement of training outcome objectives and organisational goals. Issues concerning the factors which affect transfer would seem a significant consideration for any evaluation attempt.

1.3.3 Approaches to Training

The purpose of this review is to consider the general philosophy of training design, development and delivery and those models or approaches which are most prevalent in the literature and practice. The training programme considered in Part II was conceived to meet perceived organisation needs and was based on a practical, as opposed to a theoretical, design framework. Relevant concepts and approaches are taken from the literature and retrospectively applied.

Goldstein (1989) describes training as a multifaceted phenomenon concerned with basic adult education processes requiring; the development of theories and methods to describe and specify the training needs of the organisation; the design of interventions to meet training needs; and the evaluation of interventions.

Reid and Barrington (1994) conceive training as a learning system and identify eight generalised approaches to training; learning by exposure; educational approach; problem centred approach; action learning approach; the systems approach; the analytical approach; the competencies approach; and the training process approach.

Bramley (1991) identifies the model of training which is most prevalent to be the *individual training model* which is derived from educational practice. The focus of this model is, as its name suggests, the individual learner and the process is one of encouraging individual learners to learn something which is deemed to be useful to the organisation and then expecting those individuals to find uses for their learning. Bramley cites craft apprentice training as an example of this approach to training.

Newby (1992) defines training in terms of 'the systematic cycle of training' where the focus is business operational objectives. The basic process is; (i) needs identification, (ii) priority setting; (iii) objectives setting; (iv) specification of learners; (v) design and development of training (curriculum, methods, and media); (vi) delivery of training; and (vii) reinforcement of learning. Newby places evaluation at the end of the cycle, but relates it back to each preceding stage as a judgement activity.

Camp et al (1986) define training as a sequential model and emphasise the preparation steps of training as important for the success of the training. The stages of the sequential model are; (i) data gathering/diagnosis; (ii) establish objectives; (iii) identify resources; (iv) develop curriculum; (v) plan logistics; (vi) perform training; (vii) facilitate transfer of learning; and (viii) data gathering/evaluation. Camp et al identify a ninth activity 'soliciting feedback' which links to each of the eight stages.

Jackson (1989) identifies a similar approach to training; (i) identification of training needs; (ii) analyse training needs; (iii) write training objectives; (iv) determine and develop programme content and methods; (v) conduct programme; (vi) evaluate programme; (vii) communicate programme results. Jackson identifies feedback loops from the 'evaluate programme' stage to stages (ii) to (vi) inclusive for programme adjustments.

The distinguishing factors of all these approaches are; (i) the identification of training needs in terms of operational objectives followed by; (ii) systematic processing of information through a training design and development stage through to; (iii) delivery of training to trainees. In the cases of Newby and Camp et al, a follow-up reinforcement / learning transfer facilitation stage is included. A further significant distinguishing factor is that they all position evaluation at the end of the cycle or process.

From this review, four key activities emerge; training needs analysis; training design; delivery; and training evaluation. The literature relating to the first three are reviewed here, and the last, as it is the subject of this thesis, is reviewed in chapters 2 and 3.

1.3.4 Training Needs Analysis

Ostroff and Ford (1989) argue that training needs analysis is an important step in the training process as it provides critical input to the development and evaluation of training programmes. This argument is supported by Wexley and Latham (1981), Gagné et al (1985), and Tannenbaum and Yukl (1992) who all maintain that a thorough training needs analysis should be conducted prior to the development of training.

Dalziel (1991) defines training needs as the gap which exists between the present skills and knowledge of employees and the skills and knowledge they require for effective performance. Training needs analysis, therefore, provides trainers with information on who needs training in terms of skills and knowledge, what the content of the training should be, and when the training should be delivered (Wexley, 1984; Reid and Barrington, 1994). McGehee and Thayer (1961) offer a framework for understanding the training needs analysis process which identifies three components; (i) organisational analysis, (ii) task (operations) analysis; and (iii) person analysis. Tannenbaum and Yukl (1992) endorsed this framework in their review of literature but found little empirical evidence to suggest it was actively applied.

1.3.4.1 Organisational Analysis

Organisational analysis is concerned with the study of the entire organisation and includes its objectives and available resources. The Training needs analysis identifies gaps between organisational performance and targets and identifies those areas where training offers a viable strategy for closing the gaps / meeting organisational goals.

Bramley (1991) positions training and development as a subsystem of the organisation which has inputs from and outputs to the organisation and argues that if this interaction is to result in increased organisational effectiveness, then priorities for training needs must relate to organisational goals.

Neale (1991) proposes the performance management approach to organisational analysis which consists of identifying specific and recognisable standards of work for all staff against which their performance can be assessed and their training needs established.

1.3.4.2 Task Analysis

A task analysis stems from the conclusions drawn from organisational analysis and determines the activities performed on the job and the conditions under which the jobs are done (Goldstein, 1986). The aim of the analysis is to collect and analyse information regarding the knowledge and skills necessary for effectiveness on the job.

Boydell (1977) describes task analysis as a process of examining a job in detail and Reid and Barrington (1994) identifies three types of task analysis; *comprehensive analysis* where every task required for the job is identified and expressed in terms of skills, knowledge and attitude; *key task analysis* where tasks are prioritised in order of importance before they are analysed in detail; and *problem centred analysis* where analysis is limited to a problem considered to have a training solution. The methodology for undertaking task analysis varies and are well documented in the literature.

Patrick (1992) identifies four methods; hierarchical task analysis; critical incident technique; task inventories; and trainability analyses. In each case, the goal is to break down a task into specific elements which lend themselves to the specification of training objectives.

1.3.4.3 Persons Analysis

A persons analysis determines which people in an organisation / department require training and what the types of training should be. The analysis determines the extent to which individuals are able to perform their jobs with respect to the requirements of the organisation. Bramley (1991) describes the intention as being to assess performance levels against those required in the job. The difference between performance and job requirements does not necessarily imply a training need and may be due to other factors such as organisational culture and structure, and rewards systems (Mager and Pipe, 1990). Training needs analysis provides a specification for training and importantly, with respect to training evaluation, identifies organisational, task, or people needs which are beyond the scope of training. Where undertaken, the training needs analysis provides a valuable input to the evaluation of training.

With respect to this study, no formal rigorous training needs analysis was conducted, although the need to address quality improvement through education and training was determined in general terms in an organisational sense. The recognition by senior company management as part of an overall operating review was politically sufficient to justify the development and introduction of the programme, and the urgency to respond to senior management outweighed the perceived benefit of a rigorous training needs analysis. This approach to training is congruent with the observations of many writers, most notably Goldstein (1989) who comments that training is a pragmatist activity which is largely divorced from its theory and published research.

1.3.5 Training Design and Development

Patrick (1992) identifies three main components to a fully designed training programme; (i) training content; (ii) training methods and strategies; and (iii) trainee characteristics. Within the literature, training content is determined through the training needs analysis (see section 1.3.4), however the content of the training subject of this study was determined largely by benchmarking (Fitz-enz, 1993). Benchmarking is a process of looking outside one's own arena of operation, whether that be a department, a corporation or an industry, to look at and compare the approaches of others to one's own operations. With the engineers' quality improvement training programme, the content was largely determined through external benchmarking of Pacific Rim manufacturers.

Training methods and strategies organise training content into a fully designed training programme and many approaches have been proposed in the literature (e.g. Goldstein, 1980; Reigeluth, 1983; and Gagné et al 1992), although little empirical evidence can be found in support of these approaches (Reid and Barrington, 1994). Whilst a detailed review of the literature is beyond the scope of this study, a general review of the key principles of training design is significant since, as the reader will discover, training or instructional design theory is focused on process. In chapter 2, the reader will observe that evaluation theory with respect to training and instruction is focused on outcomes. This dislocation of focus forms part of the basis for the metaevaluation in Part III of the study.

A significant amount of the training design literature is referenced as 'Instructional System Design' which is most prominent in the USA. Reigeluth (1983) defines instructional system design 'as a discipline that is concerned with understanding and improving one aspect of education: the process of instruction' and describes the purpose as 'to devise optimal methods of instruction to bring about desired changes in student knowledge and skills'.

Instructional system design literature refers to *instructors* and *instruction*, whereas other writers refer to *trainers* and *training* in a general sense, and *classroom* or *trainer-led training* in an instructional sense. For the purposes of this review, I have used the term *trainer* throughout, and *instruction* to refer to classroom or trainer led training.

1.3.5.1 Theoretical Bases for Training Design

Stammers (1987) describes both education and training as areas of the applied psychology of learning and as such they hold to a common theoretical base. Stammers observes that both can be generalised as instructional activities, where the aim is to produce some degree of control over learning events. The theory of instruction provides the applied theoretical bases for training design.

Bruner (1966) describes the theory of instruction as important in providing rules concerning the most effective way of achieving knowledge or skill. He suggests that there are four major components in a theory of instruction; (i) the experiences which most effectively implant in the learner a predisposition towards learning; (ii) the structure of the content (which he refers to as the body of knowledge) so that it can be readily understood by the learner; (iii) the most effective sequences in which to present training materials to be learned; and (iv) the nature and timing of rewards and punishments in the process of learning and instruction (or as he describes teaching).

Within the literature, many theories of training design are offered, however Reigeluth (1983) identifies eight perspectives as being significant to the design of instruction (training); *Prescriptive Model of Instruction* (Gagne and Briggs, 1979); *Behavioural Approach to Instructional Prescription* (Gropper, 1974, 1975, 1983); *Algo-heuristic theory of instruction* (Landa ,1983); *Structural Learning Theory* (Scandura, 1983; Scandura and Brainerd, 1978); *Cognitive Theory of Inquiry Teaching* (Collins and Stevens, 1983); *Component Display Theory* (Merrill,1983); *Elaboration Theory of Instruction* (Reigeluth and Stein, 1983); and *Motivational Design of Instruction* (Keller, 1983). With the exception of Gropper's (1974) *Behavioural Approach to Instructional Prescription* and Scandura and Brainerd's (1978) *Structural Learning Theory*, these perspectives are also acknowledged by Patrick (1992) as being significant to the design of training.

To avoid a lengthy discussion in the main body of this thesis, these theories are reviewed in appendix A, where their relevance to the training programme design presented in part II is also considered.

Snelbecker's (1983) review of these instructional theories relate all to; (i) the analysis and establishment of educational goals; (ii) the initial state of the learner prior to instruction; (iii) the conditions that interact with training methods and effectiveness, but cannot be directly manipulated; and (iv) feedback to the learner.

The establishment of goals and learning objectives not only provides the specification, or blue-print requirements of training, but also provides a scale against which aspects of training and training outcomes can be measured. This is not a novel idea for the evaluation of training and their importance cannot be overstated. These goals and objectives are central to training; they are derived from organisational, person-related or task related needs which encompass the potential contribution of training to the organisation; and they are written in terms of the learner and implicitly distinguish the achievements that can be derived from training and those which are beyond its scope.

The initial state of the learner encompasses a wide array of factors. The motivation of the learner, the temporal relevance of the content of the training, and the prior level of knowledge and skill are all identified within the body of theory reviewed as having significance to the design, and consequently the effectiveness, of training. From the perspective of the wider conceptualisations of evaluation (e.g. Context-Input-Process-Product; see Appendix B), the initial state of the learner is of particular interest.

Conditions that interact with training, and are external to training in the sense that they cannot be directly manipulated by the training design or delivery, are important noise factors which potentially determine to varying degrees the effectiveness of a training programme or intervention. As these noise factors are significant to training, then they have significance for the evaluation of training. Evaluation in isolation from noise factors is unlikely to yield reliable information for the improvement and decision making goals of training evaluation.

Finally, feedback to the learner is an important aspect of training design. This is not restricted to the structured learning environment, but extends to the learner's continuing learning as part of the deployment (or not) of trained skills to the workplace. Evaluation has the potential to provide feedback at varying degrees of specificity to learners long after training is completed.

From chapters 2, 7 and 8, it will become very apparent that the design of evaluation is interdependent with the nature of its evaluand and yet little attention is given to this in the literature. The theories and methods of training and the conceptualisations and methods of training evaluation are in the main treated separately.

1.3.5.2 Bloom's Taxonomy of Educational Objectives

A recurring aspect of instructional design is the use of instructional goals. Gagné et al (1992) define a goal as a desirable state of affairs and in the context of instruction provide an achievement aim for instruction. Whilst goals provide the general achievement aims for training or instruction, they are considered to be too generic and are therefore often translated into instructional or learning objectives.

With respect to the design of the programme, significant attempts were made to use learning objectives to drive the design of the programme. These provided valuable criteria for the subsequent development of level 2 evaluation instruments. The learning objectives employed were those of Bloom (1956). This selection was determined by the existing knowledge of members of the development team and whilst employed with varying interpretations of the different design and development team members, broadly followed Bloom's conception.

Bloom's taxonomy comprises general and specific categories of outcomes of instruction. The intended outcomes are expressed in terms of the intended behaviour of students - the ways in which they are able to act, think, or feel as a result of participating in training. Bloom identifies three major domains; (i) cognitive domain which is concerned largely with information and knowledge; (ii) affective domain which relates to attitudes, emotions and values; and (iii) psychomotor domain which involves muscular and motor skills. Within the programme only the cognitive and affective domain were applied, the psychomotor domain was considered not to be appropriate.

1.3.5.2.1 The Cognitive Domain

This domain is based on a set of progressive categories ranging from the knowledge of facts to the intellectual process of evaluation. Each category within the domain includes the behaviour at the lower levels. There are six major categories within this domain and these are summarised in Table 1-I.

Bloom's top three levels; analysis; synthesis; and evaluation are sometimes collectively described as 'invention' (Walkin, 1990). Learners demonstrate inventiveness by breaking down information into parts (analysis), combine with parts from other information to form new concepts (synthesis) and review those new concepts within a given context (evaluation).

Category	Description
Knowledge:	The first major category is the foundation of the domain and it is concerned with recall and methods of dealing with recalled information, such as knowledge of terminology and specific facts, knowledge of ways and means of dealing with specifics (i.e. conventions, trends and sequences, classifications and categories, criteria and methodology), and the knowledge of the universals and abstractions in a field (i.e. principles and generalisations, theories and structures).
Comprehension:	Comprehension builds on knowledge extends to the students ability to understand the meaning of information and translate it from one format to another (e.g. drawings to component dimensions). It refers to the students ability to interpret information and then to use the information to predict consequences.
Application:	This involves the student's ability to use knowledge gained in new situations. This may involve the application of principles, theories, and rules to work situations, or other learning situations.
Analysis:	This is concerned with the student's ability to separate learned material into its component parts to understand its interrelationships and organisational structure. E.g. As well as recognising the concept of quality, the student will be able to distinguish the customer, value, and continuous improvement as its component parts and how they relate to each other in an overall context.
Synthesis:	Synthesis extends from analysis in that it refers to the student's ability to combine separate elements to form a 'new' concept. This may involve deduction and other aspects of logical thought.
Evaluation:	The highest category is evaluation and is concerned with the ability of the student to judge the value of learned material against relevant criteria. This will involve challenging the validity of concepts in different contexts and knowing when something is appropriate for a given purpose.

Table 1-I: Bloom's Taxonomy: Cognitive Domain

1.3.5.2.2 The Affective Domain

The second domain is, in concept, concerned with attitudes ranging from a student's simple reception of stimuli to the complex ability of value concepts to characterise situations or behaviour. There are five major categories within the domain and these are summarised in Table 1-II.

Category	Description
Receiving:	Receiving is the basic category and is simply concerned with a students ability to receive signals from his or her environment. The student will be aware and willing to receive signals.
Responding:	As an extension of receiving, responding involves the arousal of a student's curiosity and the acceptance that the student has a responsibility in his / her response to the signal.
Valuing:	This involves the student's recognition of the intrinsic value, or worth, of a situation. Through this recognition, the student's motivation is heightened and the student's beliefs emerge from the encounter.
Organising and conceptualising:	Organising and conceptualising is concerned with the student's ability to investigate attitudes and values in a situation and to then make patterned responses on the basis of the investigation.
Characterising by value or value concept:	As the highest category in the domain, this involves the student's ability to see ideas, attitudes and beliefs within and surrounding a situation and integrate them into a coherent whole.

Table 1-II: Bloom's Taxonomy: Affective Domain

1.3.5.3 Application of behavioural objectives

During the development of the programme initial attempts were made at writing behavioural objectives within the cognitive domain for both the technical and people skills elements of the programme.

It was intended that the objectives prepared for each of the modules would serve three main purposes: i) to provide a design specification for the instructional elements of each module, ii) to be used to compare and develop the content linkages between the modules, and iii) for the simultaneous writing of the case study. Whilst development of the behavioural objectives provided the general direction and shape for the design and development of material they were found to be time consuming in the short term. As management pressure to launch the programme increased, they became less of a priority in the minds of the course designers and were not kept up to date in line with changes to the programme (Brittle, 1991).

1.3.5.4 The value of Bloom's Taxonomy.

During the process of developing the programme, I observed the desire amongst many of the development team members to "get on with writing the material as we know what we want" (Brittle, 1991).

Writing learning objectives is a time consuming task and this problem was exacerbated by the development team's lack of experience with this approach to training development. Apart from providing agreed targets, behavioural or instructional objectives provide criteria against which training can be evaluated. As training evaluation is not institutionalised in training development in the organisation considered here and elsewhere (Phillips, 1990; Goldstein, 1991), the value of learning objectives was not fully recognised or shared by the development team during the early developmental stages of the programme.

Within the literature, criticism of Bloom's taxonomy have tended to be based on three grounds:

a) Bloom's taxonomy is derived from the view of learning inherent in behaviourism. The taxonomy accepts learning as a response to stimuli, the desired response being the behavioural outcome, or objective. According to Sockett and Pring (1970), critics of the taxonomy maintain that educational objectives should not be merely behavioural, and that outcome should not be equated necessarily with learning as someone may learn without being able to convey to another evidence of having learned. Further, critics argue that behaviour assessment should not be accepted as the only reliable indicator of the attainment of those goals set by a training programme.

b) The theory of human knowledge on which it is based is naïve and inadequate. A more fundamental criticism of the taxonomy is the way it conceptualises human knowledge (Scheffler, 1961). It has been argued that 'knowledge' does not have any real significance in isolation from 'comprehension' and 'application' and therefore Bloom has ignored much of the contemporary analysis of the cognitive processes associated with epistemology.

c) Its cognitive / affective dichotomy is inaccurate. Bloom's separation of the cognitive and affective domains has been criticised as unreal (Enever and Harlen, 1972; Dembo, 1981). These critics argue that the interrelationships between the cognitive and affective dimensions of learning are such that lessons should be planned with the two integrated with each other - necessitating integral lesson objectives.

From my own experience with attempting to use Bloom's taxonomy, I found difficulty in consistently distinguishing between the various levels and found myself including behavioural objectives within the affective domain in the cognitive domain taxonomy. These difficulties were shared by my colleagues on the development team.

In spite of the theoretical inadequacies of the taxonomy and from my observations of applying the taxonomy, from a practitioner's perspective Blooms underlying concepts go a considerable way to providing an organised framework within which objectives can be stated, classified and operationalized.

1.3.6 Training Delivery

Training delivery is concerned with the interface of training with the learner. A traditional model of training delivery is that of classroom style lessons, with trainees attending a trainer facilitated programme designed to achieve stated objectives. This model of training continues to be predominant amongst practitioners. Sloman (1994) observes that the majority of training in the UK is delivered through classroom style training, with less than 10% being delivered through alternative methods, such as computer based training.

The engineers quality improvement training programme (see chapter 4) is modelled largely on this approach, however a consultant role was identified for the trainers to facilitate the application (transfer) of trained skills to the engineering process. Phillips and Shaw (1989) define a consultant as 'someone who is in a position to influence change but who has no direct authority to implement change programmes'.

1.3.7 Education and Training in an Industrial Context

Marshall and Tucker (1992) state that "the future now belongs to societies that organise themselves for learning". What Marshall and Tucker have recognised is that knowledge and skills are as important, and perhaps more so, as natural resources to economic prosperity. An exemplar of this is Japan; a nation with a fraction of the natural resources of the USA and yet with a greater Domestic Gross Product per capita. Ashton and Green (1996) argue that, as modern forms of technology are primarily knowledge based and intensive in the use of conceptual skills of the full range of the modern workforce, education and training is now of "paramount importance" in the competitive global economy.

The empirical study is centred on a training programme intended to support the host company's quality improvement objectives.

1.4 Summary

The inference to be drawn from the literature is that learning and training are inextricably linked by their nature and the role they play in an organisational sense. The theories of learning identify internal and external mechanisms to the learner and the relationships between them and with learning. By definition, these mechanisms are significant to learning and hence to training.

The theories of training advocate simplification of complex concepts for learner consumption and emphasis the need for reinforcement or practice to address longevity and transfer goals. The wider body of knowledge of training, particularly in the area of transfer, offers strategies to facilitate further learning which occurs away from the training context.

Despite these inferences to the relationships between the concepts of learning and training, writers have failed to make explicit the relationships. For example, not one of the popular training theories, is couched within a learning theory framework. As Cannon-Bowers and Tannenbaum (1991) comment the theories of learning and training are not generally integrated and research findings are not translated into useful training methods.

Chapter 2

Conceptualisations of Evaluation

2. Conceptualisations of Evaluation

> *'But spectacles have a function, and they function only when you put them on, to look through them at the world.'*
>
> *Popper, 1971*

2.1 Introduction

The term 'evaluation' is an extension of the word 'value' which was borrowed from the old French word 'value' - a noun use of the feminine past participle of 'valoire' meaning 'be worth'. Its origin is found in the Latin word 'valere' meaning 'to be strong, to be of value' (Ayto, 1991). Evaluation, therefore, is subject to interpretations of value.

This chapter is concerned with the review of the literature relating to the conceptualisations of evaluation. The conceptualisation of primary focus is that of Donald Kirkpatrick (1959; 1960) as this is the basis for the evaluation in Part II. However, evaluation is a broad ranging topic and its history and multiple interpretations provide a rich source of ideas and perspectives for the evaluation of training. This chapter is therefore organised in four main sections; scientific inquiry; conceptual diversity of evaluation; Kirkpatrick's framework; and political factors.

The general notion of scientific inquiry provides the foundations for evaluation and it is therefore necessary to briefly review this broader context. Throughout its modern history, many researchers have conceived approaches to evaluation reflecting the contemporary interpretations of value and methodology in the literature. To avoid a lengthy review in the main body of this thesis, descriptive overviews of the conceptualisations are drawn from the literature and given in appendix B. The fourth main section of this chapter considers the political factors which are significant to any evaluation, with particular emphasis on those aspects which are significant to the empirical study.

2.2 Scientific Inquiry

To understand the conceptualisations of evaluation, it is necessary to consider the wider concept of scientific inquiry and how it relates to evaluation. Heron (1981) describes scientific inquiry as a process of disciplined exploration which leads to knowledge stated in propositions. However it is at this point 'knowledge', that the literature divides into multiple interpretations of what constitutes knowledge and therefore which processes for gaining knowledge are acceptable. Whilst the purpose of this thesis is not to analyse the various interpretations and come to some siding conclusion, the significance of the differing interpretations cannot be ignored as they have relevance to both the conceptualisation of evaluation and the methods that are employed in the evaluation process.

Gross (1992) identifies two major philosophical influences on the laws which govern acceptable knowledge; (i) Empiricism; and (ii) Positivism. In exploring these influences it is necessary to comprehend the array and power of belief systems which have developed around them. For this reason, the reader will find it helpful to share my understanding of Paradigms - a concept which, in my experience, is often misunderstood.

2.2.1 Paradigms

My introduction to the notion of paradigms was during a planning meeting for the training programme of interest to this study. I watched a training video entitled "The Business of Paradigms" (Barber, 1987) in which the narrator described a scenario where a man driving his car along a winding mountain road had to swerve to avoid an oncoming car which was being driven erratically. As the oncoming car passed, the lady driver shouted "PIG!", to which the man replied "SOW!", not wanting to let the lady get away without responding to the insult - especially as it was she who was driving erratically. The man drove around the corner, feeling good that he had managed to get his insult heard when he drove into a pig which was in the road, lost control and went over the edge of the mountain.

Of course, the message here was that the man interpreted the word 'pig' as an insult and not as a warning. The man had interpreted the word through his paradigm, and not through that of the lady's. A paradigm therefore is a belief system; a filter through which we, as individuals, see the world at large. Many of us share similarities in our paradigms although sometimes, as the parable illustrates, we have quite different paradigms. Paradigms are a product of our life experiences and, therefore, they are unique. Kuhn (1970) describes a paradigm as what the members of a scientific community share, and conversely a scientific community consists of men who share a paradigm (p.176). For Kuhn, scientists work within paradigms which are general ways of seeing the world and which dictate what kind of scientific work should be done and what kinds of scientific theory are acceptable. Paradigms are deeply embedded in all of us; they tell us what is important, legitimate, and reasonable (Patton, 1990).

2.2.2 Empiricism

According to Gross (op.cit.), the empiricists were seventeenth-century British Philosophers (i.e. Locke, Hume, and Berkeley) who believed that the only source of true knowledge about the world is sensory experience; what comes to us through our senses or what can be inferred about the relationships between such sensory factors. Empiricism is one of the main discriminators between science as a source of knowledge and other sources.

The acceptance of knowledge is dependent on its source. Kerlinger (1986) and Krathwohl (1985) cite Cohen and Nagel's (1934) four sources of knowledge; (i) personal observation and experience; (ii) intuition; (iii) tradition; and (iv) authority.

(i) Krathwohl (1985) regards personal observation and experience as the source one most trusts; it is the method one uses in childhood to explore the world. For Kerlinger and Krathwohl, observation and experience is the foundation of science.

(ii) Intuition encompasses propositions that are obviously true because those propositions make obvious sense to us as individuals. The mere statement of such intuitive propositions is often sufficient for their acceptance as knowledge. However, history has repeatedly shown that intuitive propositions which are obviously true can be deceiving. A classical example of this is the intuitive proposition that the Sun revolved around the Earth. From looking at the Sun in the sky, our ancestors saw the Sun orbiting the Earth and intuitively inferred this phenomenon. For them, this made obvious sense as the Earth was created by God and as such was at the centre of the universe. It wasn't until the sixteenth century that Copernicus challenged this knowledge (Atkinson et al, 1993).

(iii) Tradition refers to knowledge which has always been true. The Bible, Koran and Talmud contain a large body of such knowledge (Krathwohl, 1985). Knowledge by tradition is passed through generations and its history provides its authority.

(iv) Authority as a source of knowledge, according to Cohen and Nagel, stems from our own limiting range of personal experience; individually we can only experience a small corner of our world. Krathwohl observes that since most propositions are not self evidently true, and since tradition tends to come with an authority figure, authorities are without question the major source of our knowledge.

As for intuitive knowledge, history has shown that tradition and authority knowledge can be untrue and hindering to the advancement of knowledge. The interesting account of traditional interpretation of the Burgess Shale fauna in 1909 by Charles Doolittle Walcott; a premier paleontologist and most powerful administrator in American science, delayed scientific discovery by 80 years (Gould, 1989). The long history of fossil interpretation, together with the authority of Walcott, meant the mis-classification of the oldest preserved soft-bodied animals with far greater potential for instruction about life's history than the dinosaurs was not challenged until long after Walcott's death.

Science is dependent on observation of actual instances (obtaining empirical data) to see how and whether something works. Within the natural sciences observation is generally facilitated by the phenomenon of interest (however it should be acknowledged that some more recent developments in the natural sciences (e.g. sub-atomic physics) have not facilitated direct observation). Within the behavioural sciences gathering empirical data on a characteristic often requires translating it into something that can be sensed. As many of the characteristics of interest are internal to human beings (i.e. learning) and they only become apparent if and when they affect observable physiological characteristics or overt behaviour. Interpreting the meaning of data therefore requires accurately inferring internal characteristics from observable characteristics. Krathwohl comments that even though behavioural sciences use empirical methods, inferences from those measures to internal states are less than certain.

2.2.3 Positivism

The term positivism can be traced back to Auguste Comte (1830) who referred to a 'positive philosophy'. According to Kolakowski (1993) , positivism stands for 'a certain philosophical attitude to human knowledge … a collection of rules and evaluative criteria referring to human knowledge'. Positivism is characterised mainly by an insistence that science can only deal with observable entities known directly by experience (Abercrombie et al, 1988). Positivism aims to construct general laws or theories which express relationships between phenomena. Observation and experience will then show that the phenomena do or do not fit the theory; explanation of the phenomena consists in showing that they are instances of the general laws or regularities.

Scriven (1996) describes the development of evaluation as having proceeded against the backdrop of the ideological battle in the philosophy of science between the positivists and their opponents, originally the idealists and later many others. Scriven argues that there are discrepancies between the positivist assertion that no evaluative judgements can be made with scientific objectivity on the evaluative judgements about the performance of science students which is made by their positivist teachers - the proponents of positivism.

2.2.4 Positivist Opponents

Positivism has been the driving philosophy behind scientific inquiry for a considerable period of time (Reason and Rowan, 1981) but there has been a major reaction to positivism in the form of phenomenology (Filmer, 1973). The phenomenological philosophy is that reality is a social construct. This implies that reality is defined by, and influenced by, social interpretations and meanings. For example, reality was once a flat world, now it is of a globe. For the scientist, the phenomenological philosophy of inquiry requires the interpretation and understanding of the social experience within a social and cultural context.

Within an evaluation context, Easterby-Smith (1994) contends that if positivism represents one pole of an inquiry methodological dimension, then constructivism represents the other. Constructivist or naturalistic inquiry (Guba and Lincoln, 1981) emphasises that there exists multiple, socially constructed realities ungoverned by any natural laws. This ontologically differs with positivism, which asserts that there exists a single reality that is independent of any observer's interest. Epistemologically, constructivists assert that an inquirer and the inquired-into are interrelated and so the findings of an investigation are literal creations of the inquiry process. With respect to this epistomological position, positivism maintains that it is possible for an observer to remain external to the phenomenon being studied (subject-object dualism). These two extreme poles are often referred to as distinct paradigms as they represent distinct and largely incompatible paradigms of inquiry (Filstead, 1979).

It is not the purpose of this review to draw conclusions which side with one approach to science (if that is ever possible). The review provides an insight into the different paradigms to give an understanding of the scope and nature of the positions. With regard to this study and the conceptualisation and approach to training evaluation, this debate about truth strongly relates to the climatic conditions within which the evaluation is conducted. If the climate is, for example, one of a positivist persuasion, then an evaluation must take account of this. If the information the evaluation yields is not framed in the language of its intended recipients, then it is unlikely the messages will be heard or understood.

2.2.5 Paradigm of an Industrial / Engineering Society

The paradigms held by those who participate in, or receive information from, an evaluation study, form an important contextual consideration for deciding on the type of evaluation to be adopted. Participants are more likely to subscribe to an evaluation which shares their rules and standards for practice as determined by their beliefs. Similarly, recipients of the information yielded by the evaluation, are more likely to be receptive to the evaluation report.

Kuhn's (1970) observation illustrates this:

> *"Men whose shared research (or learning) is based on shared paradigms are committed to the same rules and standards for scientific practice"*
>
> *Kuhn (1970) - brackets added*

The predominant scientific paradigm in industry is positivism. Engineering is grounded in this paradigm and its graduates, who constitute the overwhelming majority of the industrial community, are more likely to respond to a positivist argument. This is not to say the other paradigms of inquiry do not have a place in training evaluation or research as they do. It is merely a word of caution with regard to the conceptualisation and implementation of evaluation in an industrial context.

As a pragmatist researcher I have struggled throughout this study; on the one hand I have sought in the methodology of the natural sciences to reduce research to concise accurate reports for action; and on the other I have experienced the diversity and essence of situations. Throughout I have endured an internal positivist-constructivist battle.

2.3 Conceptual Diversity of Evaluation

Within the paradigms of inquiry exist conceptualisations of how evaluation is undertaken. A conceptualisation is a general notion; a theme or a design, of an activity (Brown, 1993). Conceptualisations of evaluation have evolved over centuries. Planned evaluation is documented as early as 2200BC with the selection of personnel in China (Shadish et al, 1995), however, this review is primarily concerned with the conceptualisations of evaluation since the 1940's, starting with Tyler, acclaimed to be the 'father of evaluation' by the Joint Committee on Standards for Educational Evaluation (1981).

2.3.1 Historical Perspective

Although emphasis and acceptance varies, there is general consensus about the evolution of evaluation and its conceptualisations in the literature (Madaus et al, 1996; Guba & Lincoln, 1989; House, 1986; Shadish et al, 1995). The general trends are discussed by Guba & Lincoln (1989), who classify the evolution into four generations; i) *measurement*, characterised by the identification and measurement of the variable (outputs) of the evaluand; ii) *description*, characterised by description of patterns of strengths and weaknesses with respect to certain stated objectives; iii) *judgement*, characterised by efforts to reach judgements about a programme, and in which the evaluator assumed the role of judge; and iv) *responsive constructive*, characterised by programme stakeholders determining the direction, scope and purpose of an evaluation.

Whilst Guba and Lincoln's general trends are useful in providing milestones of the general shifts in emphasis of evaluation, they are set in the context of responsive-constructivist evaluation (see appendix B: Fourth Generation Evaluation). Madaus et al (1996) offer a potentially less biased account of the history of evaluation, drawing on the perspectives of a wide range of researchers. Madaus et al distinguish six periods in the evolution of evaluation; age of reform (prior to 1900); age of efficiency and testing (1900-1930); Tylerian age (1930-1945); age of innocence (1946-1957); age of expansion (1958-1972); and age of professionalization (1973 to date). The ages of reform and of efficiency and testing are concerned with the early development of testing in schools and the development to standardised achievement tests and norm referencing. However it was not until the Tylerian age when any significant developments in the conceptualisation of evaluation occurred (Madaus et al, 1996).

In the 1940s, Tyler conceptualised evaluation as the comparison of intended outcomes with actual outcomes. This was significant as prior to this, evaluation had been conceptualised as quantitative and experimental; a process of comparative research using experimental and control groups. Tyler's conceptualisation allowed internal comparisons of a programme's outcomes with its objectives, thereby avoiding the need for control groups. Within education, Tyler's approach was used to help define objectives for the USA High School curricula and to assess the degree to which the objectives were realised. Purposes for evaluation were largely concerned with the comparison of performance of educational establishments.

From 1945 to 1957, no significant developments were made in how evaluation was conceptualised. Madaus et al. comment that the post-war prosperity meant there was little call for educators to demonstrate efficiency and effectiveness. During this period however there were developments in some of the technical aspects of evaluation, most notably the development of educational objectives (Bloom, 1956) for the design of instruction and for its evaluation. The period was characterised by the emphasis on outcomes in terms of intended objectives and elaboration of the types of outcomes. This coincided with the publication of Kirkpatrick's articles on evaluating training (see section 2.4).

The age of expansion was triggered by the Russian launch of Sputnik in 1957 which led to new large-scale educational programmes in the USA, necessitating evaluation of these efforts.

Cronbach (1965) conceptualised evaluation as a process of gathering and reporting information that could help guide programme development. In reviewing the Tylerian educational evaluations, Cronbach observed that by analysing and reporting test item scores, as opposed to average total scores, evaluators could identify areas of a programme which needed improvement. Cronbach extended the purpose of evaluation to include improvement.

Through the 60's conceptualisations of evaluation along the theme conceived by Tyler and Cronbach emerged (i.e. Discrepancy model (Provus, 1971) - see appendix B), but more significantly, Stufflebeam (1966) and Stake (1967) offered a radically different conceptualisation of evaluation. They extended the scope of evaluation by considering the goals, inputs, implementation and delivery, and measurement of the intended and unintended outcomes of a programme. They also raised the issue of whether a potential evaluand was worth evaluating. Stufflebeam's CIPP model of evaluation is featured in appendix B.

Shadish et al observe that during the 1970s several journals dedicated to evaluation were started against a backdrop of fragmented evaluation work. They describe this as the start of the age of professionalization and since then, there has been significant development in the conceptualisation of evaluation. Parlett and Hamilton (1972) suggested evaluation as an illuminative process (see appendix B) where critical issues of an evaluand emerge through a flexible observation and interview methodology that capitalises on available resources and opportunities. Scriven (1974) is credited with introducing the concept of goal free evaluation (Easterby-Smith, 1994) which is based on the belief that the only way an evaluation can avoid being contaminated by those with vested interests in the evaluand is for the evaluation design to take no account of the formal goals and objectives of a programme. Scriven argues that this is essential for a balanced judgement about the real value of a programme.

Stake (1975) conceptualised evaluation as a responsive process (see appendix B) whereby the evaluation is conducted with no predetermined design and is concerned with responding to events that are noticed about the evaluand and are of interest to evaluand stakeholders. Eisner (1976) conceived evaluation in terms of art critic (appendix B) where the evaluator is an expert in the evaluand's subject matter and who provides criticism in the form of description, interpretation and appraisal of his experiences with the evaluand.

Guba and Lincoln (1989), drawing largely on the thinking of Scriven and Stake, conceived evaluation as an emergent process which is dependent on inputs from stakeholders. Guba and Lincoln offer a twelve stage process for conducting evaluation (see appendix B).

The inferences to be drawn from the historical perspective of evaluation with respect to this study are;

(i) There is no integrative theory of evaluation; evaluation has evolved against the backdrop of the philosophical debate within the social sciences (Scriven, 1996) with the result of divergent conceptualisations of evaluation. Easterby-Smith (1994) suggests that these can be viewed on scientific / constructivist dimension (Figure 2-I). Morrison (1993) offers a similar analysis, referring to quantitative / qualitative continuums.

Scientific		Constructivist
Measuring things (Quantitative methods)		Qualitative Methods
Absolute Criteria	⟷	Multiple Values
Preordinate		Discovery
Evaluator Controls		Shared Control
Observer Separate		Observer Involved

Figure 2-I Easterby-Smith's Scientific / Constructivist Dimension

(ii) Implicit in the literature is evaluation as methodology of an evaluator; a person or persons, external to the evaluand and distinct from evaluand stakeholders.

(iii) The development of evaluation has remained largely distinct from that of its evaluand; evaluation within the context of education and training is not integrated with theories of learning and training.

(iv) Central to any conceptualisation of evaluation is the judgement of value; much of the debate concerning how evaluation is conceptualised has focused on how value is judged and the criteria for judgement.

2.3.2 Evaluation Models

Within the general conceptualisations of evaluation exist abstract frameworks, systems or approaches. These offer ways in which evaluations are undertaken and have been described as 'evaluation models' in the literature (Scriven, 1991), however Stake (1981) disputed the term 'model', preferring to call them 'persuasions'. Models are abstract simplified descriptions of phenomena or systems (Brown, 1993).

Scriven (1991) describes models as being to paradigms as hypotheses are to theories. The models of evaluation are less general in their application than the paradigm and conceptualisation from which they are derived as they are specific to their range of application contexts.

The significant models of evaluation have previously been drawn from the literature as part of the review of its history. However, within the context of the evaluation of training, and with particular relevance to this study, the most notable and widely accepted evaluation model (Cascio, 1987; Carnevalle & Schulz, 1990; Gordon, 1997) is that of Donald Kirkpatrick (1959, 1960). This model is the basis of the training evaluation programme described in Part II of this thesis and therefore I have elected to describe and critique it in some detail (Section 2.4).

2.3.3 Training Evaluation

Training (including education) is a multi-million dollar business (Patrick, 1992) and yet many writers have stated how few and how inadequate attempts to evaluate training have been. Phillips (1990) cites a survey of management training conducted in 1977 where 24% of executives (sample size: 3100) reported that changes in job behaviour was measured.

Settle (1987) observed that the comparison between the need for evaluation and the amount of evaluation that occurs represents one of the most significant shortfalls in education.

2.3.3.1 A Definition of Training Evaluation

Defining training evaluation is difficult. Ralph Tyler (1950: pp.69) perceived evaluation as:

> *"The process of determining to what extent the educational objectives are actually being realised"*

Whilst Tyler was primarily concerned with juvenile education and so his definition is set in an educational, and not training, context, his definition ties evaluation to the educational objectives of the program. The definition relates evaluation to the objectives, or outcomes, of the evaluand.

Patrick (1992) offers a wider definition of evaluation:

> *'Evaluation is any attempt to obtain information concerning the effect or value of training in order to make decisions about any aspect of the training programme, the persons that have been trained and the organisations (local, national or international) responsible for providing that training.'*

However, Patrick himself acknowledges that this definition is too narrow to cover the evaluation approach proposed by Warr, Bird and Rackham (1978) which encompasses training needs analysis and the inputs to decisions to be made as to which type of (training) intervention should be used - both of which are considered before the training programme has been designed, let alone implemented.

Goldstein (1986) offers a broader definition of training evaluation which encompasses a range of purposes and avoids the specific reference to the effect or outcomes of training:-

> *'... the systematic collection of descriptive and judgemental information necessary to make effective training decisions related to the selection, adoption, value, and modification of various instructional activities.'*

Along similar lines, Brandenberg and Smith (1991) take a wider view of training evaluation, and introduce the notion of client as the provider of the purpose of the evaluation:-

> *'Evaluation is a judgement of an entity on some dimension valued by the client. This judgement is based upon a measurement of actual status on the dimension against a standard.'*

In using the term judgement, Brandenberg and Smith are referring to an appraisal or a decision which is generally divided into two classes: summative and formative. Summative evaluations are attribute decisions; they are directed at go / no go decisions (see chapter 3 for review). Audiences for summative decisions, note Brandenberg and Smith, are often senior management of the client organisation. In this context, Brandenberg and Smith view the evaluator as external to the organisation, with evaluation being done to the organisation.

Formative evaluations seek to identify ways of improving the evaluated entity. Brandenberg and Smith note that trainers are generally the audience for formative evaluations because they are the people who manage and improve the course. Summative and formative evaluations can be supported by the same study.

By using the term 'entity', Brandenberg and Smith are defining evaluation in a wider context than just a training program. Entity may also refer to any aspect of the training or Human Resource Development, trainees, instructors, instructional strategies, facilities and the training organisation itself. Other organisational interventions such as feedback and incentive systems, team-building, personnel selection and placement can be included.

The term Dimension refers to the 'critical characteristic' of the entity that is valued by the audience for the evaluation.

"Measurement of actual status" is defined as collecting data to show how 'things really are' (or at least how people think they are) in terms of the evaluation dimension. Measurement techniques may include opinion data gathered by interview or questionnaire, knowledge testing, performance testing under simulated job conditions, observation of job performance, organisational measures of job performance.

A 'standard' is a criterion for judging success of failure. Standards may be relative (norm-referenced) or absolute (criterion referenced). Relative standards may take the form of pre and post entity comparisons of performance. Absolute standards are determined independently of the evaluation entity, for example, using job performance criteria, production measures.

Despite the broad ranging definitions of evaluation expressed in the literature, the empirical study adopted a well established and narrow definition. This was largely determined by the framework which was to be employed, namely Kirkpatrick. Kirkpatrick (1959a, 1959b, 1960a, 1960b, 1994) defined evaluation in terms of the outcomes of training. He does not give a concise definition of training, arguing that to clarify the 'elusive term' evaluation requires an elaborate explanation, hence his four level model.

2.4 Kirkpatrick's Model for Evaluation

In 1959 and 1960, Donald L. Kirkpatrick (1959a, 1959b, 1960a, 1960b) published a series of articles in the Journal of the American Society for Training (ASTD). The articles described four steps of evaluation that he had originally conceptualised in his PhD dissertation at the University of Wisconsin, Madison, USA. He developed his four step model, or taxonomy, in an attempt to clarify what he describes as the elusive term evaluation (Kirkpatrick, 1994).

Since Kirkpatrick's original articles, many training evaluators have adopted his ideas. The American Society for Training and Development (ASTD, 1990) reported that the evaluation framework that most training practitioners use is the Kirkpatrick model and that it is the most widely known evaluation model. Other writers have commented similarly, regarding Kirkpatrick's framework as; (i) the predominant framework for the evaluation of training (i.e Shennan and Lockhead, 1996); (ii) having significantly shaped the human resource development profession (i.e. Gordon, 1997); and (iii) having become part of the language of training evaluation in many large commercial organisations (i.e. Basarab and Root, 1996). Cascio (1987) observed that the field of industrial / organisation psychology has largely accepted the framework of Kirkpatrick for the evaluation of training and Brown and Seidner (1998) describe Kirkpatrick's framework as the 'de facto' model for the evaluation of corporate training.

Kirkpatrick initially described his thoughts in terms of four steps, although these are more often referred to as Kirkpatrick's four levels (Goldstein, 1986; Alliger and Janak, 1989; Kirkpatrick, 1996 and 1998). Kirkpatrick's four steps have consequently been examined by Kirkpatrick and others and relationships between the steps have been inferred. For the purposes of this review, the four steps are first considered separately and then the relationships, or concept of framework, is reviewed.

2.4.1 The Four Steps

Kirkpatrick's four steps are; reaction (Step 1); learning (Step 2); behaviour (Step 3); and results (Step 4). Interpretations of the meaning and purpose of these levels have remained largely unchanged since their conception.

2.4.1.1 Step 1: Reaction

According to Kirkpatrick (1959a; 1994), evaluation on this level measures the reaction of participants to the training programme; how much the trainees liked the course or program. Kirkpatrick likens it to measuring the feelings of the trainees and Alliger and Janak (1989) have conceived it as a measure of trainee attitudes towards the training. Kirkpatrick identifies some standards for collecting reactions data including; use of a written comments sheet which can be tabulated and quantified; obtaining honest reactions by making the forms anonymous; and providing trainees with the opportunity to write additional comments not covered by the questions.

Kirkpatrick describes reactions as the first step in the evaluation process and regards them as important. Determining how people feel about the programmes they attend is an indicator of customer satisfaction (Kirkpatrick, 1996) and decisions by senior management within an organisation are frequently made on the basis of one or two comments they receive from people who have attended (Kirkpatrick, 1996; Easterby-Smith, 1994; Basarab, 1994). For Kirkpatrick, the systematic collection of documented trainee reactions to training ensure that such decisions are made on the feedback from a representative sample population. Kidder and Rouiller (1997) comment similarly, observing that documented positive trainee reactions help ensure organisational support for the training programme.

Kirkpatrick made explicit in his original article that 'even though he (the training manager) has done a masterful job of measuring the reaction of the group, he still has no assurance that any learning has taken place. Neither that has he any indication that the behaviour of the participants will change because of the training programme'. Kirkpatrick (1959b) re-emphasised this in his introduction to level 2 evaluation saying 'it is important to recognise that favourable reaction to the programme does not assure learning'. Kirkpatrick is clear in this sense that he implies no relationship between level 1 and the subsequent levels of his framework. The significance of this statement becomes clearer, when I examine the subsequent interpretations of Kirkpatrick's framework in the literature.

Reactions evaluation is widely regarded in the literature to be easy to apply; Alliger and Janak (1989) observe that reactions to training can be measured during the training seminar using standardised questionnaires as do others (i.e. Patrick, 1992; Shennan and Lockhead, 1996). Kirkpatrick also notes that as reaction is so easy to measure, nearly all training directors do it, although not necessarily to the standards he identifies (Kirkpatrick, 1994).

2.4.1.2 Step 2: Learning

For the purposes of his original discussion of training evaluation, Kirkpatrick defined learning as:

> *"What principles, facts, and techniques were understood and absorbed by the conferees"*

This interpretation of learning within the context of Kirkpatrick has remained unchanged; Basarab and Root (1992) conceive learning in these terms and as the basis for reactions methodology design, as do Kemp (1995), Kidder and Janice (1997) and others. This aspect of Kirkpatrick's evaluation framework is also accepted by his critics (i.e. Holton, 1996).

The reader should note that Kirkpatrick's definition is not consistent with more widely accepted definitions of learning offered by Kimble (1961); Coon (1983); Atkinson et al (1993) in that it does not include a temporal dimension; that is to say the concept of permanency does not feature in his definition. This enabled Kirkpatrick to distinguish 'learning' from 'behaviour', which he considered in step 3. The reader may find it useful to think of Kirkpatrick's learning as knowledge gain, when reading the remainder of this section.

To measure learning, Kirkpatrick offers guidelines (termed 'guideposts') for establishing a procedure for measuring the amount of learning that takes place during a training program. The guidelines recommend measuring individual performance in terms of before and after course knowledge / skill / attitude levels, measuring learning objectively and, where possible, using a control group (not receiving the training). He also advocates analysing the results using statistical methods to prove learning in terms of correlation or level of confidence.

Methods suggested are organised into 'Classroom performance' for programmes concerning job instruction (i.e. work simplification, interviewing skills, reading improvement, effective speaking and effective writing) and 'Paper and Pencil Tests' where principles and facts are taught, as opposed to techniques. Classroom performance methods were those concerned with classroom activities performed by the trainees and observed by the trainer, or instructor. Kirkpatrick recommended that these form an integral part of the program.

Pencil and Paper tests were questionnaires administered before and after the course. Analysis was recommended in terms of two approaches: a) the total score of each trainee, and b) the responses to each item (test question).

Kirkpatrick advocates a high degree of planning for level 2 evaluation to ensure that data collected could be easily analysed and interpreted. Purposes given for conducting level 2 evaluation were similar to those for collecting level 1 information; to provide the training manager with objective data to use in selling future programmes to the organisation and for increasing his status and position in the company.

Alliger and Janak (1989) observed from their extensive review of the literature, that reactions and learning are often measured using a single instrument which accommodates the designs for both purposes. Basarab and Root (1992) suggest similar tests to those of Kirkpatrick.

2.4.1.3 Behaviour

Measuring behaviour, according to Kirkpatrick (1960a; 1994; 1996) is concerned with the extent to which the knowledge / skills and attitudes taught in the program are transferred to the job. Kirkpatrick acknowledged the difficulty of transfer of training and its measurement. He advocates, what he described as "a more scientific approach", which takes account of factors other than the training program itself. Kirkpatrick cited nine training evaluation studies which he considered to be best practice (step 3) studies at the time of his original article, which are summarised in Table 2-I to explain 'Behaviour' evaluation.

All of the studies shared the characteristic of on-the-job performance assessment, and many used before and after comparisons and experimental and control groups. Information was not only collected from the trainee, but also from the trainees' work colleagues. All of the examples cited by Kirkpatrick were of management or supervisory training programmes.

Kirkpatrick's advocation of control and experimental groups to isolate the affects of other factors on the measured behaviours due to training is, in the context of contemporary literature, clinically naïve with respect to programmes which are of an organisational change scale. It takes no account of Bandura's social learning theory, or Sengé et al's concept of learning organisation. Regarding the empirical study in Part II, where the programme spans a minimum of 6 months, it is unlikely there will be no effects on the behaviour of the control group due to either of these types of learning.

Name of Study	Program	Study Design
Fleishman-Harris Study (1955)	Central School of International Harvester Training Program	7 post-course paper and pencil questionnaires. Information obtained from trainees, supervisors, and subordinates
	Central School of International Harvester Training Program	Before and after measure of job performance using experimental and control groups. Information obtained from trainees and their subordinates
Survey Research Center	Human Relations Program at the Detroit-Edison Company	Before and after measure of on-the-job performance with experimental and control groups. Information obtained from trainees and subordinates.
The Lindholme Study (1950 - 51)	Insurance company training program	Before and after training questionnaire. No control group used. Information obtained from trainees' subordinates.
The Blocker Study (1955)	Democratic Leadership training course in a large insurance company (600 employees)	Post training measure of trainee behaviour through a study of structured records of interviews with employees kept by the trainees (the trainees did not know the records were to be used for behaviour evaluation).
The Tarnopol Approach (1957)	Supervisor training program.	Pre and post course employee attitude surveys including neutral questions not relating to the training, using experimental and control groups. Information collected from trainees' subordinates
The Moon-Hariton Study (1958)	Manager Training Program at General Electric Company	Attitude and behaviour questionnaire (administered 2 years after the program) identifying and quantifying changes since the program Information collected from trainees and subordinates.
Buchanan-Brunstetter Study (1959)	Supervisor training program at the Republic Aviation Corporation.	Post-course on-the-job behaviour questionnaire describing behaviour and identifying "items more effectively done now than a year ago" administered to experimental and control group. Information collected from trainee subordinates.
The Shroud Study (1959)	Personal factors in Management training program at the Bell Telephone company, Pennsylvania.	Post-course 3 part questionnaire consisting of behaviour scale questions and Critical-Incident questions and course objective achievement questions administered to experimental and control groups. Information collected from trainees and their supervisors.
The Sorensen Study (1958)	Crotonville Advanced Management course of the General Electric Company	self, subordinate, peer and supervisor observation of post course behaviour using experimental and control groups. Information collected from trainee, peer group, subordinates, and supervisors.

Table 2-I: Kirkpatrick's (1960) Exemplar Evaluation Studies: Behaviour

Kirkpatrick summarised evaluation of behaviour as being essential to training managers and their programmes as, in his opinion, the effectiveness of training in outcome behavioural terms determined the future of training in an organisation.

More recently, Kirkpatrick (1994) cites nine examples of implementation of his framework (Table 2-II), which are characteristic of his earlier citings in respect of measurement of behaviour. Although the methods of data collection have become sophisticated, the general interpretation and design of step 3 evaluation are similar in respect of measurement against training objectives and the use of pre/post training comparisons.

Name of Study	Programme	Study Design
First Union National Bank Study (O'Hara, 1994)	Communication, awareness, renewal and empowerment training programme for nonexempt bank employees	Combined learning and behaviour questionnaire sent to random sample of trainees 1 to 3 months after training.
Kemper National Insurance Study (Clarke, 1994)	Performance appraisal and coaching training for staff managers and supervisors (pilot course)	Performance appraisal report form checklist administered to all trainees and performance appraisal questionnaire administered to all staff of trainees whose performance had been appraised.
Intel Study (Freitag, 1994)	Three day senior management programme on; systems thinking; managing fast-cycle organisation; and maximising performance.	Post training review of degree of implementation of action plans established during training. Data contrasted with training needs analysis data to establish pre/post training behaviour differences.
Motorola Study (Basarab, 1994)	Three day creative manager training programme for business managers	Post training self assessment questionnaire sent to initial 50 trainees and questionnaire sent to 3 work associates identified by trainee.
St Luke's Study (Wagner and Weigland, 1991)	Outdoor-based communication, trust and empowerment training programme for new hospital managers. Three 1 day sessions.	Self assessment behaviour questionnaire administered to all participants before and after training. Post training questionnaire administered twice; 3 month and 6 months after training.
Arthur Andersen Study (Bond, 1994)	Two 1 day session effective presentation skills training for organisational consultants.	Pre and post training observations using behaviourally anchored rating scales (and recorded on video-tape). Further observation at Follow-up session (1 to 8 months after initial training) .
Automotive Industry Study (Holcombe, 1994)	Two-week sales induction training for new district managers.	Observation of trainee behaviour using gap-analysis course objectives based checklist by trainee's manager.
Wisconsin Study (Kirkpatrick, 1994)	Three day supervisory skills training programme: public course.	Pre training questionnaire (all trainees) and post training structured interview conducted with 60% of trainees. Post training interview with trainee's manager.
CIGNA Study (Paquet et al, 1994)	Six I day session management skills training programme for all managers	Pre/post Likert scale survey questionnaire measuring seven indices of management skills administered to random sample of trainees.

Table 2-II: Kirkpatrick's (1994) Exemplar Evaluation Studies: Behaviour

2.4.1.4 Results

In his original article (Kirkpatrick, 1960b), Kirkpatrick classified results of a training program in terms of; reduction of costs, reduction of turnover and absenteeism; reduction of grievances; increase in quality and quantity of production; or improved morale. Kirkpatrick acknowledged the difficulty of measuring certain programs in terms of results because of the many complicating factors (other than the training programme) which also affect results by citing Keachie:

> *"Difficulties in the evaluation of training are evident at the outset in the problem technically called 'the separation of variables'; that is, how much of the improvement is due to training as compared to other factors?"*

Kirkpatrick recommended that evaluators begin to evaluate in terms of Reactions, Learning and Behaviour. It is unclear from the article whether Kirkpatrick is suggesting that steps 1 to 3 are precedents to evaluating results, or if, because of the inherent difficulty in results evaluation, he is recommending that efforts are best placed with steps 1 -3 and not with step 4.

To describe step 4 evaluation, Kirkpatrick offers 6 case studies of evaluation at the results level. These are summarised in Table 2-III.

Name of Study	Program	Study Design
Beekman Study (1958)	Plant Safety Training: Colgate and Palmolive.	Comparison of Pre and post training plant safety records.
Massey Study (1957)	Accident Prevention training: US Mail	Experimental group received 35 hours of training; control group received no training. Accidents recorded by category; comparison of experimental and control group accident incidence by category.
Schallert Study (no date)	Insurance estimation and appraisal of automobile damage training: Farmer Mutual Insurance	Insurance adjustors attended a 3 week training programme. Comparison of trained adjustors' logs of estimates (6 months) to garage's repair estimates in 6 months after training.
Wisconsin Cost Reduction Institute Study (1958)	Cost reduction training for production supervision: Wisconsin University open course	Depth interviews with trainees and trainee managers to establish estimated cost savings. Post training questionnaires sent to un-interviewed trainees to establish estimated costs.
Merrihue and Katzell Study (1955).	Occasional study: General Electric	Employee Relation Index comprising employee relations concerns (i.e. absenteeism) using data drawn from employee records.

Table 2-III: Kirkpatrick's (1960) Exemplar Evaluation Studies: Results

Concluding the article, Kirkpatrick observed that evaluation of training programs in terms of results "is progressing at a very slow rate", but adds that " in years to come, we will see more efforts along this direction".

Kirkpatrick's step four has been further elaborated by Hamblin (1974). Hamblin devised a 5 level framework, or taxonomy, of training evaluation. He based his framework on Kirkpatrick's 4 steps, however he distinguished results in terms of organisational variables (level 4) and ultimate value variables (level 5). Organisational variables included non-economic factors, such as productivity, quality, and employee morale. Ultimate value variables included economic factors, such as sales volumes, costs and profits, similar to Kirkpatrick.

Advocacy of results criteria has become increasingly important in the literature in the last decade (i.e. Phillips, 1991; Kemp, 1995) and shifts in emphasis within human resource functions to be perceived as value added as opposed resource draining (Hall and Goodale, 1986; Walker, 1992; and Ashton and Green, 1996) have strengthened attitudes towards step 4 evaluation within the training profession (Phillips, 1994 and 1998).

2.4.2 Relationships between the Steps

In his original series of articles, Kirkpatrick described the four steps separately, inferring no causal relationship or hierarchy between the steps, other than the numerical sequence assigned to the steps. His only acknowledgement of order was in terms of the increasing difficulty of each step; evaluation using reactions was easier than evaluation of learning, which was less difficult than evaluation of Behaviour, with evaluation of results being the most difficult of his 4 steps.

Subsequent writers have inferred relationships between the steps. Hamblin (1974) suggested that the levels formed a causal hierarchy: good reactions lead to learning; learning leads to improvements in job behaviour; improvements in job behaviour lead to improvements in organisational variables (improved quality, increased productivity, improved morale); and improvements in ultimate value variables (increased sales, reduced cost, improved profits). In Hamblin's view, for evaluation at any given level to be meaningful, evaluation of the preceding level in the hierarchy had to be undertaken. Noe and Schmitt (1996) suggested a similar causality hierarchy.

Clement (1981) hypothesised that causal relationships existed between Hamblin's levels 1, 2, 3 and 4: Positive reactions correlated to learning achievement; learning achievement correlated to improvements in on the job behaviour; and improvements in on the job behaviour correlated to improvements in organisational variables. Due to the difficulty in measuring ultimate value variables, the relationship between organisational variables and ultimate value variables was not considered. To test his hypotheses, he conducted a correlational study using a 1st line supervisor training course. A pre-test / post-test control group design was employed and measures were developed for the 4 levels. From his study, Clement found only partial support for the hierarchy of training outcomes: Reactions were positively related to learning, but learning was only positively related to one of the three training course behavioural objectives or outcomes. Clement concluded that Hamblin's hierarchical model "did not reveal the inference of variables beyond the training course which can influence the outcomes of training".

2.4.2.1 Variables affecting the relationship between Reactions and Learning.

Clement identified three "especially important" variables affecting the relationship between Reactions and Learning: the trainee's readiness for the course; the trainee's motivation to take the course; and the opportunities for practice and feedback during the course.

2.4.2.2 Variables influencing the relationship between Learning and Job Behaviour

The two primary variables influencing the relationship between learning and job behaviour were identified by Clement to be: lack of similarity between the training and job settings; and a lack of opportunity to apply the training to the job.

2.4.2.3 Variables influencing the relationship between job behaviour and Organisational Variables

Variables identified by Clement which influenced the relationship between job behaviour and Organisational Variables are less precise than those given above, however he identifies two categories: a) within the organisation variables, and b) outside the organisation variables. Within the organisation variables includes factors such as: the trainees manager; the trainee's peers; and organisational policies. Outside the organisation variables refer to the influence of environmental forces, such as government relations and competition.

In addition to Hamblin's assumption of a causal relationship between Kirkpatrick's steps, Alliger and Janak (1989) identified two further assumptions which they perceived to "appear to be largely implicit in the minds of researchers and trainers" about Kirkpatrick's taxonomy. Firstly that Kirkpatrick's steps are arranged in ascending value of information provided (Newstrom, 1978); and secondly, that these steps are positively inter-correlated, that is to say that apart from a causal relationship, there is a correlation in positive outcomes between the levels. This is a subtle distinction from Hamblin's assumption in that Hamblin suggested that favourable outcomes at one level lead to (cause) favourable outcomes at the next level. Alliger and Janak made an extensive review of the literature (1960 - 1988) to examine the validity of these assumptions. They found limited evidence which supported and rejected all these assumptions, concluding that such assumptions were problematic.

Kirkpatrick's response to these interpretations has been contradictory; in his initial articles he stated that there were no relations between the steps; in 1994, he suggests there are relationships between the levels - favourable outcomes at one level lead to favourable outcomes at the next (Kirkpatrick, 1994); and more recently (Kirkpatrick, 1996), he states 'I don't care whether my work is called a model or a taxonomy as long as it helps to clarify the meaning of evaluation in simple terms'. This final statement probably best reflects Kirkpatrick's pragmatism to the subject of training evaluation.

For the purposes of this study, Kirkpatrick is used as an organising framework for the evaluation and, as such, merely provides a taxonomy of training outcomes to provide a focus for the design of the evaluation, as originally conceived by Kirkpatrick (1959a; 1959b; 1960a; 1960b).

2.4.3 Critics of Kirkpatrick's Framework

Kirkpatrick has been widely accepted amongst training practitioners and researchers and as such direct critics of his framework are rare. Holton (1996) criticised Kirkpatrick's framework on two counts; (i) that it is not integrative in the sense of the relationships between the levels; and (ii) that relationships between the levels are not testable.

Alliger and Janak (1989) observe that not all training is intended to affect change at all levels; training may be largely rewarding; spirit building; or as a prerequisite to another course of training. In this sense step 4 evaluation would be inappropriate. Kirkpatrick's framework becomes more of a taxonomy or checklist for evaluation, as opposed to a systematic organising framework.

Phillips (1998) identifies several problems with Kirkpatrick's framework. In addition to those given above, Phillips argues that as the framework does not isolate the effects of training (important to levels 3 and 4), training will improperly take credit for any improvement. From the earlier reviews of Kirkpatrick, this problem has arisen because of the way in which the framework has been applied (Table 2-II and Table 2-III), as opposed to how evaluation has been defined. In this sense, Phillips criticism is of implementation and not of the framework itself.

2.4.4 Application of Kirkpatrick to Training Evaluation

Following publication of Kirkpatrick's series of articles (1959 and 1960), the field of industrial/ organisational psychology largely accepted this framework (Cascio, 1987; Carnevalle & Schulz, 1990). And as mentioned previously, Kirkpatrick's framework was used as the basis of development of further frameworks. Industry and Commerce at large also adopted the framework and it has appeared in numerous texts on training published in the last 30 years and is well known among those involved in the development and delivery of training (Plant and Ryan, 1994). In almost every general text on training I have reviewed, Kirkpatrick's framework has featured. The company employ Kirkpatrick's framework as part of their corporate evaluation strategy (FMC, 1993). Training professionals use Kirkpatrick's terms to describe the type of evaluation that will be employed for a given program.

This is, perhaps, not surprising; the simplicity and broad ranging scope of Kirkpatrick to outcomes facilitates inclusion of any evaluation study which considers outcomes of training in Kirkpatrick's terms. Of the twelve exemplars cited by Kirkpatrick (1994), only three evaluated training using all four steps; one study using three steps; four studies using two and four studies using just one of the steps. Where studies did not use all steps, these were not confined to the lower order steps; three of the one step studies were at level 4.

Furthermore, Alliger and Janak's (1989) study illustrates the apparent difficulty in applying all 4 levels to training evaluation. Of the 203 published evaluation studies, only 22% were undertaken at 2 levels, 3% at 3 levels and less than 2% were undertaken at all 4 levels.

2.4.5 Contribution to Training Evaluation

It is important that we do not lose sight of the contribution that Kirkpatrick has made. He provided a crude taxonomy of training outcomes, within which training evaluation activity can be categorised. Despite its popularity in the literature and in the language of the training professionals, so-called full Kirkpatrick framework evaluation studies are rare. Alliger and Janak's (1989) extensive review of the literature of published 'Kirkpatrick framework' studies between 1969 and 1989 revealed that of the 201 studies identified, only 3 studies where evaluation had been conducted at all four levels; 5 studies at three levels; 44 studies at two levels; and 149 studies at one of the levels. Ralphs and Stephan (1986) reported that 86% of the top 500 USA companies evaluated training courses only at level 1. The Department of Employment Training Agency (1989) reported that of the 80% (by workforce) of UK companies surveyed, 90% used level 1 evaluation but less than 10% attempted level 3 evaluation and only 3% attempted to evaluate the results of training on the business (level 4).

In the view of the writer, in outlining a simple taxonomy of training outcomes, Kirkpatrick has provided the basis for an evaluation paradigm; a framework for evaluation which focuses solely on the outcomes of training and to the exclusion of other factors of training, which are often the areas of training which programme managers can do something about (Weiss, 1986). In this sense, Kirkpatrick has defined evaluation for many contemporary training practitioners and established a set of implicit rules governing evaluation of training which are widely accepted within the training and research profession. Kirkpatrick's framework is concerned only with the outcomes of training and gives no attention to issues of operation. The implications of this are considered in Part III of this study.

2.5 Political Factors

Much of the literature accepts evaluation is greatly influenced by politics. Politics is a term which encapsulates the various interests of persons involved or affected by an evaluation. Morrison (1993) describes evaluation as being prone to internal and external biases which are inescapable because they involve people. By internal biases, Morrison is referring to the way interested parties affect the design and construction of the evaluation; and external biases refer to the way in which information derived from an evaluation is used.

Political factors vary considerably across different organisations. Easterby-Smith (1994) observes that most organisations have unique value systems which are shared by large numbers of employees working for them. Individuals, or departments align themselves behind one value system or another and this is reflected both in the way they attempt to influence decisions and in how they interpret information that is of relevance to those decisions. Competing interpretations of what is right and appropriate is also influenced by social or personal values.

Rossi and Freeman (1993) reflect the view of many other writers (i.e. Easterby-Smith, 1986 and 1994; Patrick, 1992; and Kirkpatrick, 1975 and 1994) that evaluation fulfils a decision making purpose and observe that decision-making is a highly complex process of which evaluation information is just one element. They contend that decision making in an organisational context is sensitive to weighing, assessing, and balancing the conflicting claims and interests of a number of constituencies. Furthermore, Rossi and Freeman argue that as evaluations take time, especially those directed at assessing programme impact, they often lag behind the political and evaluand time scales which often move much faster.

The evaluation, therefore must either be designed so as to provide information on a timely basis with regard to the needs of the programme or the politics, rather than yield data which, in a political sense, is invalidated for its intended purpose as it arrives long after decisions have been made, and sometimes after they have been implemented.

It should be noted that Rossi and Freeman hold to a formative-summative paradigm of evaluation. Like many other writers cited in this review (i.e. Patrick, 1992; Reid and Barrington, 1993; and Basarab and Root, 1995), evaluation is viewed as either formative or summative. Rossi and Freeman's argument reflects the summative view of evaluation, where evaluand results or outcomes constitute the evaluation information, and therefore time-lag is inevitable. Real-time evaluation data, that is to say receiving information on the same day or in the same hour, is not a feature of evaluation literature in an educational or training sense. This issue is explored further in chapter 7 where parallels are drawn with evaluation in other sectors of society, albeit under a different descriptor; that of 'quality management'.

From the literature it is clear that evaluation cannot be independent of the politics which surround the evaluand. Patton (1978) observes that many social scientists want to be non-political in their evaluation research, however, to be innocent of the political nature of evaluation is to wittingly, or unwittingly, become a pawn in someone else's game.

2.6 Summary

The literature with respect to evaluation falls into two distinct categories; research and pragmatic. The research literature features primarily in the development of evaluation and has resulted in multiple conceptualisations of evaluation. With the pragmatic arena of training, evaluation has tended to remain firmly grounded in commercial value criteria, where debate with respect to value has been confined largely within a framework of commercial objectives; whether they are concerned with organisational change or with contribution to a company's prosperity in financial terms.

Kirkpatrick's model is the most widely used evaluation approach within training. It was conceived in the early modern history of evaluation and its underlying conceptual principles of measurement of outcomes in terms of predetermined goals have remained unchanged. In many quarters (Basarab and Root, 1994), Kirkpatrick's model has become integrated into the language of training practitioners and in this sense its underlying conceptualisation has not been challenged. Developments in the wider social arena of evaluation have had little influence on Kirkpatrick and its proponents, however practices born out of developments in this wider context have been employed within the Kirkpatrick model, e.g. the notion of stakeholders and their needs with respect to evaluation information.

2.6.1 Research Focusing

From the reviews of the literature with respect to learning, training (chapter 1) and the conceptualisation of evaluation in the context of this study opportunity, two research focusing questions emerge:

1) What role does training evaluation, conceived in terms of Kirkpatrick, play in training improvement within a commercial context?

2) What restructuring is necessary for training evaluation to integrate with the theories and practice of learning and training?

The purpose of these questions is to centre the research on the key issues of evaluation role and integration, but not to the exclusion of other important evaluation issues which may emerge through the empirical study. As stated in the introduction (page 1), this research deliberately sets out to consider training evaluation in the broadest sense and not be confined by established and deep rooted beliefs, or paradigms.

Chapter 3

Measurement and Methodology

3. Measurement and Methodology

'Measurement is a process to understanding.'

Deming, 1994

3.1 Introduction

Across the range of conceptualisations of evaluation presented in the literature, measurement provides a universal way of describing, comparing and valuing phenomena. Evaluation methodology, the methods of evaluation vary across conceptualisations, but they provide the means by which data is collected and analysed. The purpose of this chapter is to examine the foundations of measurement, provide a general overview of evaluation methodologies, and review those methods which are employed in this study.

3.2 Measurement

The development of the natural sciences was facilitated by a comprehensive system of measurement which is universally accepted. Kyberg (1984) describes measurement as being fundamental to the physical sciences and to engineering; scientific results are judged on the basis of evidence, and convincing evidence can only be provided by measurement.

In comparison to the natural sciences, Kerlinger (1986) describes measurement in psychology and education as being misunderstood. He observes that measurement with respect to the natural sciences is largely intuitive, however measurement of characteristics of individuals and groups is much harder to understand even though it shares the same thinking and general procedure.

The purpose of measurement, whether applied to the natural or social sciences, is to provide a valid, trustworthy, traceable representation of chosen entities or phenomena of which selected attributes are of interest. Measurement can be used to describe an existing object or phenomenon, or a future system by using measurement to describe its specification.

Kaposi and Myers (1994) offer the following definition of measurement:

> *Measurement is the process of making empirical observations about referent entities of the world, and representing their properties in a formal symbol system, so as to describe them.*

The term 'referent entities' is used to describe tangible items (i.e. parts of cars, or items of clothing) or abstract notions (i.e. learning, or temperature). The referent entities may be single items or a class of entities. A formal symbol system is an organised set of symbols which represent values of referent entity variables (i.e. the weight of a car part expressed as kilogrammes). A measure is the product of the measurement process and therefore a measure is a symbol of a symbol system, designating the value of a property of the referent. The organisation of the symbol system is determined using a model.

Where the measurement of a property assigns a value directly to the referent attribute, it is termed *direct measurement* (e.g. weighing a casting, where the referent attribute is mass, the property variable of weight can be directly measured from the measurement scale). Where it is not feasible or convenient to obtain the property measure of the required attribute by direct observation, property measures of other attributes related to the desired attribute in a known way can be taken. From the related property measures, the property measure of the desired attribute can be inferred. This type of measurement is termed *indirect measurement* (e.g. to infer the density (referent attribute) of a casting, related property variables of volume and mass can be measured, from which the density can be calculated (inferred).

3.2.1 Modelling

To cope with the complexity of social systems, or an evaluand such as the engineers' quality improvement training programme, it is necessary to replace it with a simpler 'model'. The training programme in this study is highly complex with many internal and external factors and the variety of relationships which exist between them. Ross-Ashby (1956) illustrates the need for models in describing a simple pendulum; 'every material object contains no less than an infinity of variables and therefore possible systems. The real pendulum, for instance, has not only length and position: it also has mass, temperature, electric conductivity, crystaline structure, chemical impurities, some radio activity, velocity, reflecting power, tensile strength, a surface film of moisture, bacterial contamination, an optical absorption, elasticity, shape, specific gravity, and so on and so on. ...... what is necessary is that we should pick out and study the facts that are relevant to some main interest ...'

Compared to the training programme considered in this study, or to any social or educational programme, a pendulum is relatively simple and yet, as Ross-Ashby illustrates, in can be described in terms of many attributes. In attempting to study an evaluand and establish measuring systems, it is necessary to model those characteristics, or variables, which are of interest.

The purpose of models, therefore, is to selectively represent the important features of phenomenon in the simplest form which meets the needs of the enquiry for which the model is conceived. Kaposi and Myers (1994) define models as 'purposefully simplified representations of one entity by another'. The original entity of interest is called the referent and its representation is the model. The simplification is achieved by selectivity; attributes of the referent which are essential to the given purposes are preserved; others are deliberately suppressed'.

Measurement, therefore, can be described as a two-step process; (i) from referent to model; and (ii) from model to measure (Figure 3-I).

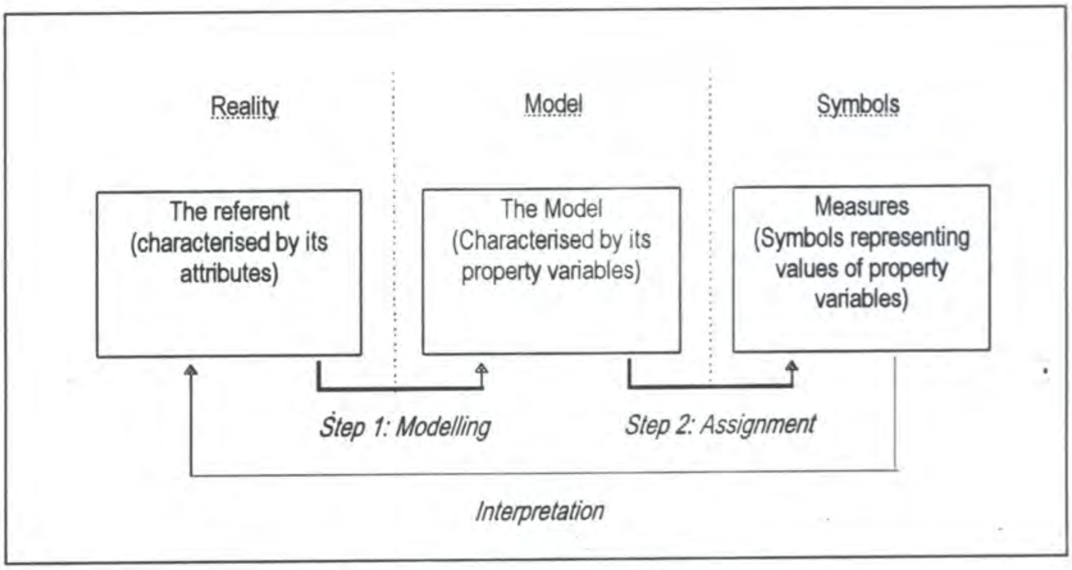

Figure 3-I: Two step approach to Measurement

In this sense, a model is an abstraction which represents selected attributes of the referent and provides their context and meaning. The chosen attributes must be defined in property variables of the model. The model also represents the inter-relationships of the attributes, thus allowing derivation of properties and characterisation of attributes which are difficult to observe directly. It is the role of the model to assure that each attribute is resolved into well defined, formally related, directly measurable properties.

From the model, a formal relation system, or measuring system, can be devised. Symbols are assigned which represent the values of the property variables. The measuring system provides an interpretation of reality (e.g. to measure knowledge gain as a result of training, knowledge, the referent attributes, are defined in terms of learning objectives. The set of learning objectives provide a model of reality with selected attributes of interest and suppressed attributes which are not of interest with respect to training. A multiple forced choice questionnaire designed to examine each learning objective is used and together with the results provides an interpretation of the knowledge gain of a particular trainee or a group of trainees; the reality).

The measurement system can be designed in a number of ways. The two general categories are quantitative and qualitative measures, which can be further distinguished into; nominal; ordinal; interval; ratio; and absolute measures.

3.2.2 Quantitative and Qualitative Measures

A distinction between quantitative and qualitative measurement features largely in the literature of evaluation of training (i.e. Goldstein, 1989; Easterby-Smith, 1994; and Harris and Bell, 1994). Much is written on qualitative and quantitative research and research methods in social investigation (i.e. Breakwell et al, 1988; Hammersley, 1993; and Cohen and Manion, 1994), however the literature suffers from a lack of simplistic description and discrimination of qualitative and quantitative measurement, preferring to discuss the two in terms of methodology. To describe and distinguish these two types of measures, the following descriptions are offered.

Quantitative measures have both magnitude and dimension. The magnitude of a measure represents the property by a specific symbol of the formal relation system of measurement. The dimension qualifies the magnitude of the quantitative measure by reference to the unit of measurement.

Qualitative measures have magnitude only. They are self contained as they incorporate all of the required information about the value of the property. Qualitative measures are dimensionless.

Both quantitative and qualitative measures provide descriptions of phenomena and have a role in the evaluation of training. Quantitative measures, by being referenced to a universally accepted measurement system, benefit from both the developments of that measurement system (i.e. statistical theory, and generalised methodology) and the common (wider) understanding of the system (i.e. it provides a common language between the evaluator and the target audience for the evaluation information). Qualitative measures, however, tend to be developed in a narrower field of activity or context and often the nature of the measure has to be communicated along with the information the evaluator is attempting to convey. Both quantitative and qualitative measures allow for comparison and value judgement and have a role in the evaluation of training.

3.2.3 Measurement Scales

Kerlinger (1986) distinguishes four types of measurement scales: nominal; ordinal; interval; ratio. Kaposi and Myers distinguish a fifth, as does MacRae (1994); absolute scale, and further distinguishes these five main types as qualitative scales (nominal and ordinal) and quantitative scales (interval, ratio and absolute).

3.2.3.1 Nominal Scale

The simplest form of scale is the nominal scale of measurement. It is a qualitative measure which classifies items of the referent by the chosen attribute. The requirements of nominal measurement are described by Kerlinger as being simple; all members of a set are assigned the same numeral (symbol) and no two sets are assigned the same numeral (symbol). The only demand of nominal scale is that that like is distinguished from unlike. Nominal scales are therefore suitable for the classification of items.

This classification or ordering of phenomena into groups on the basis of their relationships is described by Sneath and Sokal (1973) as numerical taxonomy.

3.2.3.2 Ordinal Scale

The ordinal scale is a qualitative measure which not only classifies items of the referent by the chosen attribute, but also to order the members of a group according to the extent to which they possess the chosen attribute. Ordinal scales impose an appropriate ordering relation over the symbols. Ordinal scales are suitable for the ordering, sorting and grading of items.

3.2.3.3 Interval Scale

The interval scale of measurement is quantitative in that it enables the magnitude of the attribute to be expressed numerically, as a distance from a chosen point of reference. Interval scales are suitable for expressing relative velocity and elapsed time.

3.2.3.4 Ratio Scale

The quantitative ratio scale expresses the magnitude of the measure as a multiple of a chosen unit of measure. A ratio relationship exists between scale values and therefore the formal relation system includes sum/difference and multiplication/division. Ratio scales are suitable for expressing distance, calendar time, IQ scores.

3.2.3.5 Absolute Scale

The quantitative absolute scale is reserved for counting and uses the rational numbers as their unique symbol system of measurement. The scale shares all the characteristics of nominal, ordinal, interval and ratio scales, but has an absolute or natural zero that has empirical meaning. The unit of measurement is deemed to be non-negotiable in that the attribute is resolvable to absolute atoms and the magnitude of the measure expresses the number of these. Absolute scales are suitable for counting.

3.2.4 Measurability

Measurability is the degree to which a characteristic of a phenomena or entity of interest (referent attribute) can be reliably and usefully measured. Reliability is considered later in this chapter. The degree of measurability is characterised by Kaposi and Myers in terms of; definitional; modelling and representational; practical; and quality requirements.

The definitional requirement concerns the extent to which the attribute to be measured is understood. Empirically observed attributes must be identified with an already defined concept, or composition of concepts.

The Modelling and representation requirement determines that each attribute must be modelled by a well defined property variable and variable values are represented in a suitable symbol system. A valid theory must be available for modelling each indirectly measured property as a structure of properties for which a procedure of direct measurement exists.

The practical requirement applies to direct measurement. A feasible procedure must exist, together with appropriate equipment, skills and other resources, for observing and recording measures of the property. Finally, the quality requirement is that the measure must be relevant to the requirements of the original problem and consistent with the needs of the users of the measure.

Measurability is an important determinant in the development of the evaluation of training. Misconceptions of measurement, particularly with regard to the manipulation of data generated by different measurement scales (as cautioned by Kerlinger) will lead to ill conceived results and conclusions in particular evaluations and provide a false basis for the development of evaluation of training technology. Kaposi and Myers' requirements of a measurability provide a safe guard against such misconceptions.

3.2.5 Validity and Reliability

Any measurement method in social research should yield both reliable and valid results (Jackson, 1991). For training evaluation these are important considerations if the feedback received from a study is to be of any true value to the stakeholders.

3.2.5.1 Validity

Phillips (1991) emphasises the significance of validity of data collection methods stating 'probably the most important characteristic of an evaluation instrument is validity'. Oppenheim (1992) offers a simple definition of validity as the degree to which an instrument measures what it is supposed or intended to measure. Dane (1990) describes validity as the extent to which a claim or conclusion is based on sound logic.

Validity can be described in four different contexts for which Phillips (1991) identifies approaches to determining whether an instrument is valid:

Content Validity refers to the extent to which the instrument represents content of the programme. Low content validity indicates that the instrument does not represent a true sample of what was covered. High content validity means that the instrument represents a good balance of all the information presented on the course.

By focusing on the information, as opposed to the course objectives, Phillips has assumed that the course content is reflective of the course objectives. The instrument should therefore be established against the programme (instructional or learning) objectives as the information presented may not necessarily meet the objectives that the course designer set out to achieve.

Construct validity refers to the extent to which an instrument represents the construct it is intended to measure. A construct is an abstract variable such as skill, attitude, ability or knowledge.

Concurrent validity is the extent to which an instrument agrees with the results of other instruments administered at approximately the same time to measure the same characteristics.

Predictive validity is the extent to which an instrument can predict future behaviours or results. This has considerable potential but does not currently apply to the pre and post knowledge questionnaires instrument.

3.2.5.2 Reliability

Reliability is the degree to which an instrument gives approximately the same results from subsequent measures of an item. Fluctuations in results are caused by errors. These can include, for example; variations in the conditions under which the instrument is administered; fluctuation in the mental alertness of the participant; differences in interpreting the result from the instrument; and random effects caused by the motivation of the participants.

As part of the development of the pre and post knowledge questionnaires developed for the evaluation of the engineers' quality improvement training programme (chapter 5), where data are collected at the beginning and at the end of the module, it is essential that the instrument is reliable otherwise the changes in scores can not be attributed only to the training.

Phillips (1991) summarises the literature and identifies four procedures which can help insure that an instrument is reliable:

Test/retest; involves administering the same test or survey to the same group of employees at two different time periods and calculating the correlation of the scores. If there is a high degree of positive correlation, then the test is reliable.

Alternate-form method; involves constructing two similar instruments and administering those to employees at the same time and analysing the correlation between the two scores. If there is a high positive correlation, then the instrument is considered to be reliable. Constructing a similar instrument is time consuming, which may make this approach impractical. Dane (1990) identifies alternate forms to overcome reliability problems due to practice effects, rapid changes in the characteristic being measured or, extended memory for previous responses.

Split-half procedure; involves splitting the instrument into two equal parts and comparing results. For example, it might be appropriate to compare the even-numbered question with the odd-numbered questions. The scores of the two halves are compared, and their correlation's are checked. Once again, a high correlation indicates a reliable instrument.

Inter-item Correlation's; A fourth procedure to measure reliability is to calculate correlation's between each of the items on the instrument. For example, a test with 25 items is divided into 25 parts. A correlation is developed comparing each item with all of the other.

With respect to this study and the field of evaluation of training in general, measurement is the foundation of evaluation. Within the natural sciences, measurement has developed considerably in the quest for knowledge. Measurement serves some decision problem, when the decision maker has to judge the referent on the basis of key properties, and the properties of interest may or may not be directly measurable. A model-based measurement scheme allows the deduction of each property of interest from those directly measurable; it then permits the object-oriented characterisation of the referent as a coherent assembly of (directly and indirectly measured) properties; finally it explicitly states the values judgements of the decision maker in utility measures.

In chapter 5, the evaluation measures used as part of this study are discussed and attention is paid to the reliability and validity of the evaluation measures. Both qualitative and quantitative measures are used in the evaluation of the training programme. The predominant scientific paradigm within the company is that of positivism, or a persuasion to the natural sciences; the core business of the company is engineering and its people, in the main, are drawn from a natural science oriented education system. This facilitates the communication of quantitative measures as the language of the measurement itself; that is to say the required knowledge to interpret the value property, is understood and widely accepted by the target audience.

Communication using qualitative measures is more difficult as the information to interpret the value properties must also be included in the communication. Necessarily, qualitative reports are far bulkier than quantitative reports and are less likely to be as easily accepted, or understood as quantitative reports.

3.3 Evaluation Methodology

A large selection of methods for the collection and analysis of data exist within the literature and the majority of these are generalised beyond the evaluation of training. The methods described and reviewed in this section are those most referred to by writers in the field of educational and training evaluation (i.e. Phillips, 1991; Newby, 1992; Morrison, 1993; Easterby-Smith, 1994; and Cohen and Mannion, 1994). The purpose of this section is to provide a general overview of the array of methodology available for evaluation of training, drawing particular attention to those methods employed in the empirical study.

Morrison (1993) describes evaluation methodology by using a series of continua (Figure 3-II).

1. Numbers and statistics	*through to*	transcripts of conversations and interviews, words.
2. Closed questions, multiple choice questions	*through to*	open-ended questions.
3. Desire to measure responses, compare one set of responses to another, to correlate responses	*through to*	a desire to capture the uniqueness of a particular situation person or programme or what makes it similar to and different from others.
4. A desire for formality and the precision of numbers and prescribed categories of response where it is known in advanced what is being looked for	*through to*	a more responsive, informal intent where what is looked for is far less predetermined to the point when it will only be known when it is found.
5. Portraying regularities of behaviour, of scores, of opinions in order to begin to make generalisations from results to describe what is happening	*through to*	portraying uniqueness, the complexity of a situation where we are trying to understand and why individuals behave in certain ways - to explain rather than to describe.

Figure 3-II: Morrison's Evaluation Methodology Continuums

Morrison's descriptions provide an appropriate introduction to this topic by illustrating the diverse nature of the methodologies of evaluation. Evaluation methodologies are require to meet a range of measurement needs and any evaluation study will potentially comprise of a combination of methods used independently or in conjunction with other methods.

3.3.1 Purposes

Evaluation methods are often regarded as the key part of any evaluation activity (Easterby-Smith, 1994), but they are not an end in themselves (Patton, 1981; Hamblin, 1974; and Guba and Lincoln, 1989); they serve specific purposes within an overall evaluation strategy and those purposes will vary depending on the type of study which is being conducted. The purpose of evaluation methods, therefore, is to gather data which are to be used as a basis for inference, interpretation, for explanation and prediction (Cohen and Mannion, 1994).

Within this specific context of purpose, a dimension which features in the literature is that of formative and summative purposes of evaluation. Although this aspect is included here as part of the review of evaluation methodology, it could legitimately be described in terms of a general conceptualisation of evaluation.

3.3.1.1 Formative/Summative Evaluation

A notion prevalent in the literature is that of formative and summative evaluation (i.e. Bramley, 1991; Stufflebeam, 1996; Scriven, 1967 and 1995; Camp et al, 1986). Scriven (1967) first made explicit the distinction of formative and summative evaluation, proposing that formative evaluation was concerned with improving the programme and summative evaluation concerned with judging its worth. Subsequently several writers have popularised the terms. Smith (1981) noted Scriven's 1967 article as being the most cited with respect to educational evaluation.

Basarab and Root (1992) interpret formative evaluation as providing information to training staff for purposes of improvement during development and implementation. They describe its basic purpose as measuring progress and to use this information for programme improvement during the life of the programme. Basarab and Root conceptualise formative evaluation in terms of Kirkpatrick's levels 1 (trainee reactions) and 2 (trainees' learning, in Kirkpatrick's terms of learning).

With regard to summative evaluation, Basarab and Root's interpretation is one of providing information to show the merit and worth of a training programme, with the basic purpose being to provide a summary report of the training results. Basarab and Root conceptualise summative evaluation as Kirkpatrick's levels 3 (trainee behaviour) and 4 (training results).

Formative evaluation has therefore been largely interpreted as a planning instrument; to be conducted during the initial stages of a training programme to identify ways in which it can be improved. This often takes the form of a pilot programme evaluation. Summative evaluation has been interpreted as an end of programme activity, often in the form of an end of programme report describing the outcome value of the training.

3.3.2 Data Collection Methods

As previously noted, a multitude of data collection methods can be drawn from the literature. These include qualitative and quantitative data collection methods, methods for use by an evaluator, methods for use by a learner, or methods which can be used by both together. Some writers have chosen to offer simple lists of methods (e.g. Rae, 1991), whereas others have linked types of data collection instruments to Kirkpatrick-type levels of evaluation (e.g. Bramley, 1991). Basarab and Root (1994) distinguish four general categories of evaluation instruments; questionnaires; interviews; tests; and observation forms, and provide adaptations in terms of Kirkpatrick.

Morrison (1993) offers a more general arrangement by, in addition to his continuums, further organising data gathering methods into *written forms* and *interpersonal forms*. Written forms include questionnaires, written tests, and action plans. Interpersonal forms include interview, practical tests, observation, and evaluation meetings.

Easterby-Smith (1994) classifies data collection methods in a way similar to that of Morrison, however he makes explicit reference to the levels of control over the type and nature of the data which is collected by a particular method. Easterby-Smith classifies ten data collection methods using a 'researcher / evaluator control' to 'subject / informant control' continuum. Easterby-Smith's array of data collection methods include those which he considers to be either; commonly used and well known; or less well known but have considerable utility for evaluators. Figure 3-III summarises Easterby-Smith's evaluator / informant classification of data collection methods.

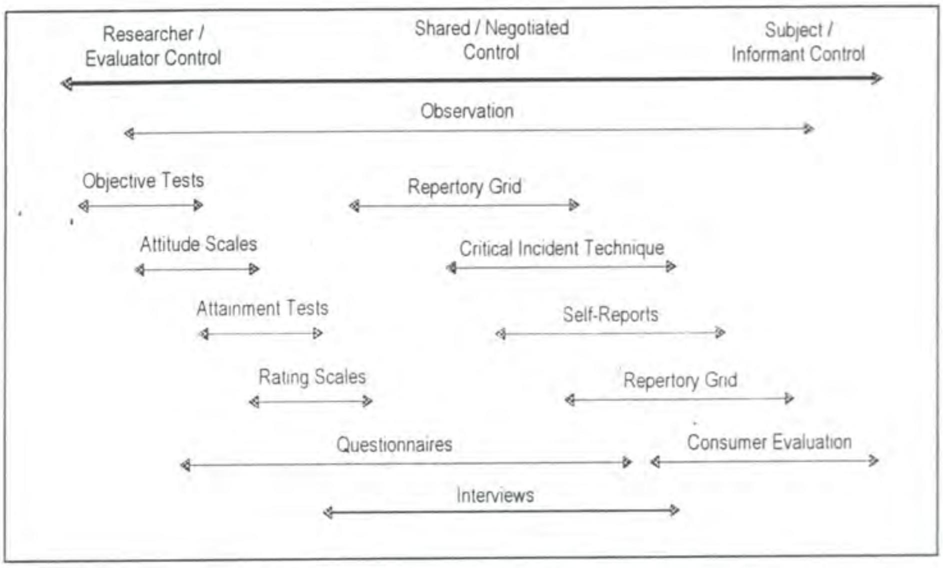

Figure 3-III: Easterby-Smith's (1994) Method Classification[1]

With respect to this study, the evaluation employs a variety of questionnaire, interview, observation and focus group methods (Table 3-I) and these are used as part of formative and summative studies.

Phenomena	Formative	Summative
Course / materials design	• Observation • Questionnaire	
Participant Reactions	• Focus Group	
Participant Knowledge		• Knowledge Gain Test
Participant Attitudes		• Attitude Survey Questionnaire
Participant Behaviours		• Critical Incidence Technique • Observation
Programme Effectiveness		• Stakeholder Analysis • Interviews

Table 3-I: Programme Evaluation Methods

3.3.2.1 Observation

Observation, perhaps more-so than any other method of training evaluation, is surrounded by the positivist / post-positivist debate and these grounds are inescapable in reviewing observation. Easterby-Smith (1994) distinguishes holistic and selective approaches to observation; holistic observation is where the observer does not apply any initial focusing or filtering and attempts to

[1] Note: Observation added.

record everything that occurs; and selective observation is where predetermined frameworks are used to guide the observation.

Guba and Lincoln (1981; 1989) contend that observation is a responsive and holistic process: responsive in that the observer must first make sense of the dimensions of a context and then seek to make those dimensions explicit; and holistic in that the observer views the world as a continuous context within which programme participants view themselves and their lives as real, true and having meaning. Guba and Lincoln argue that to make sense of the world, the observer must immerse him/herself in the environment and suspend his/her own value judgements.

Kerlinger (1986), whose persuasion is towards a positivistic approach to the behavioural sciences, identifies two main concerns with observation from his perspective; (i) relating observed behaviour to constructs or variables of interest; and (ii) the effect of the observer on the subjects of observation. With respect to his first concern, one of interpretation of behaviour by an observer, Kerlinger argues that the greater the burden of interpretation, the greater the validity and reliability problems, therefore making it necessary to define what is to be observed 'fairly precisely and unambiguously' (p. 489) by providing the observer with an operational definition of the variable being measured in behavioural terms.

Kerlinger distinguishes molecular and molar approaches to categorising behaviour to facilitate observation and inference; molecular approaches take smaller segments of behaviour (e.g. instances of saying 'I' and 'we' in a team setting) as units of observation; whereas molar approaches take larger behavioural wholes (i.e. instances of 'accepting others' approaches, suggestions, and ideas' in a team setting) as units of observation. The balance between the two is one of reliability and validity; molecular approaches are generally easier to use with less scope for observer error (high reliability), but reduce behaviour so that it no longer bears resemblance to the behaviour it is intended to observe (low validity); whereas molar approaches use broad natural definitions achieving a high degree of validity, but are more open to interpretation by observers (low reliability). Kerlinger suggests that molar observers interpret the meaning of behaviour based on experience and knowledge and molecular observers seek to push their own experience, knowledge and interpretation out of the observation process.

Another aspect of relating observed behaviour to constructs of interest is concerned with sampling and Kerlinger identifies two types; event sampling and time sampling. Event sampling is the selection for observation of integral (as part of the wider context) behavioural occurrences or events in a given class and the observation process must occur when the events take place. Time sampling is the selection of behavioural units for observation at different points in time, which can be systematically or randomly selected. Kerlinger argues that event samples possess continuity of behaviour in lifelike situations and as such possess inherent validity, time samples increase the probability of obtaining representative samples of frequently occurring behaviour, but suffer from lack of continuity and adequate context.

With respect to the effect of the observer on the subjects of observation, Kerlinger asserts that subjects will quickly adapt to an observer's presence and act as they usually act, providing observers are 'unobtrusive and not give the people observed the feeling that judgements are being made'. Patton (1982) comments similarly noting that participant observation can be more or less obtrusive depending on the nature of the situation, the kind of data to be collected, the skill of the observer, and a variety of other factors.

Patton's (1990) treatment of this range of views on observation is to distinguish observation in terms of; observer levels of involvement; use of overt or covert observation; duration of the observation; and focus of the observation. Patton's distinctions are necessarily complex (i.e. each is subject to many factors or characteristics of the evaluand; and each is subject to interpretation from a range of philosophical, ethical and practical perspectives) and therefore he describes them in terms of continuums. These are summarised in Figure 3-IV.

	Level of Involvement	
Full Participant Observer	Partial	Onlooker Observation as an outsider
	Overt / Covert	
Overt: all evaluand staff know observation is being made; by whom; and for what purpose	Some evaluand staff are aware/ partial knowledge of purpose	Covert: evaluand staff are not aware of observation (deceived)
	Duration	
Single limited duration observation		Long-term, multiple observations
	Focus	
Narrow: Single element or component of evaluand	Partial Observation	Broad: holistic view of entire evaluand

Figure 3-IV: Patton's (1990) Observation Continuums

For the purposes of this study, observation was used as part of the pilot programme evaluation and as part of the Kirkpatrick level 3 evaluation. Pilot programme observations were conducted by members of the development / delivery teams, who were participant observers. Their role was primarily to deliver the training, however as part of this process, trainee behaviours were observed. No predetermined frameworks or criteria were established and observations were essentially based on the tacit knowledge of the observers and the purpose of the pilot study; to assess the feasibility of the design, content and process of the training.

The observation for the implementation evaluation was more rigorous. Behaviour observation rating scales (see section 3.3.2.6) were developed from trainer knowledge of the people skills content of the programme using critical incident technique (see section 3.3.2.2).

3.3.2.2 Critical Incident Technique

Developed by Flanagan (1954), critical incidents are defined as "extreme behaviour, either outstandlingly effective or ineffective with respect to attaining the general aims of the activity" (p.338) and the principle of the technique is to focus on these to describe the activity. In this way it is intended to throw light on the realities of normal behaviour and circumstances (Ruddock, 1981).

Patrick (1992) suggests using the technique to define tasks required to perform a job, and in this sense the technique is used as part of training needs analysis, and Easterby-Smith suggests using the technique to identify changes which have occurred as a result of training. Critical incidents data can be collected using observation (as advocated by Flanagan), or by questionnaires or interviews.

In the context of this study, critical incident technique was used as part of the development of behaviour observation rating scales (section 3.3.2.6), where focus group interview method was used to elicit critical incident data from programme trainers with respect to people skills. .

3.3.2.3 Interview

Bingham and Moore (1959) defined the research interview as a conversation with a purpose. Oppenheim (1992) describes the purpose of evaluation interviews as being "to obtain information of certain kinds .... in the form of factual replies to factual questions, or responses to attitude scale items, or ideas and feelings, or perceptions and expectations, attitudes and the like" (p.66).

Evaluators are concerned primarily with two types of interview; (i) exploratory interviews; and (ii) standardised interviews (Patton, 1990). The purpose of exploratory interviews is primarily to develop ideas. Exploratory interviews are generally free-style interviews, and can include group interviews or focus groups (Krueger, 1988), and depth interviews (Ruddock, 1981). The purpose of standardised interviews is primarily that of quantitative data collection. Standardised interviews, sometimes referred to as structured or closed interviews are essentially verbal structured questionnaires (Patton, 1990).

For the purposes of this study, a group interview was employed as part of the pilot programme. This was conducted to explore a range of aspects of the programme and to understand the participant's perspectives on how it could be improved. In this sense, the interview was also a depth interview.

Group Interviews, sometimes referred to as focus groups, are conducted with a selection of interviewees who are asked to reflect on questions asked by the interviewer, either verbally or written. Developed largely by Merton et al (1956), group or focus interviews serve the purpose of obtaining high quality data where people can consider their own views in the context of others. Brown et al (1988) described group interviews as "not just a convenient way to accumulate the knowledge of individuals ... but give rise synergistically to insights and solutions that would not come about without them" (p.40).

The main advantages of group interviews to be drawn from the literature are: (i) they are relatively efficient for qualitative data collection; (ii) larger sample sizes can be used than individual interviews with the same interviewer resources; (iii) the interview process provides some quality control on the data collection as participants provide checks on each other; (iv) the groups dynamics contribute to focusing on the most important topics; and (v) it is easier to assess the extent to which there is consistency of views of participants.

The main disadvantages are; (i) the amount of response time to each question to facilitate group input; (ii) the need for skilled process facilitators to conduct the interviews; (iii) the need for note-taking simultaneously with interview facilitation often requires two interviewers[2]; (iv) that unexpected diversions are likely to occur in the group; and (v) it is not possible to guarantee confidentiality.

Depth interviews (Ruddock, 1981) are intended for collecting information about how people think and feel about the topics of concern to the evaluator. As the primary objective is to maintain respondent spontaneity, Ruddock suggests the interview should consist of a continuous monologue by the respondent with little input from the interviewer. This requires skilled interviewers as they must note not only what is said, but what is not said by noticing hesitations and exploring what lies behind them. Because depth interviews generate large amounts of data and are intended to explore perceptions and attitudes, they should be recorded using audio tape which can

[2] The use of Metaplan techniques negate the need for two facilitators - see Chapter 5: Pilot Evaluation.

subsequently be transcribed and analysed. For the pilot evaluation, interviewee comments were recorded using written process, known colloquially as 'metaplan'. A description of this process, together with its application is given in chapter 5.

The literature suggests that the main advantages of depth interviews are (i) that they provide a rich source of data and explore beyond general statements to examine attitudes and perceptions. This has particular significance for the evaluation of training as it enables specific insights into participants' (and stakeholders' in general) attitudes and perceptions of the management and effectiveness of training and training organisations; (ii) the interviewee can be assured of confidentiality; and (iii) differences in perceptions between interviewees can be explored in detail.

The main disadvantages are; (i) the relative high costs in terms of time and interviewing expertise; and (ii) the relatively low sample sizes which are feasible, with large programmes making the data statistically invalid.

3.3.2.4 Stakeholder Analysis

Stakeholder analysis features in the literature both as a data collection tool (i.e. Guba and Lincoln, 1989; Burgoyne, 1992) and as a conceptualisation of evaluation (i.e. Weiss, 1986). For this reason it is included both here in its context as a data collection method, and in chapter 2 and appendix B as a more general conceptualisation of evaluation. Burgoyne (1992) describes stakeholder analysis as a data collection method which recognises the interested parties who affect, or are affected, by the evaluand.

The stakeholder approach can be used to define an evaluation study in advance, but for the purposes of this study, stakeholder analysis is used to collect evaluation data from stakeholders who have experienced the training.

A stakeholder is anyone who affects or is affected by the program. Guba & Lincoln (1981) define a stakeholding audience as a group of persons having some common characteristics that has some stake in the performance (or outcome or impact) of the evaluand, that is somehow involved in or affected by the entity being evaluated. By virtue of holding a stake, an audience has a right to be consulted about its concerns and issues, to have those concerns and issues honoured by the evaluator as he goes about his tasks, and to receive reports from the evaluator that are responsive to those concerns and issues. The evaluator, in turn, has the right to prioritise the audiences in terms of the level of stake each holds, and to respond to them in that priority order to the extent that his resources permit.

This definition expands that previously offered by introducing the rights of a stakeholder and the responsibility that is held to them by the evaluator. The definition also introduces the notion of prioritising stakeholder audiences or groups.

For the study, stakeholder groups were identified as; technical vice presidents and senior managers; programme development team (course designers); programme trainers; programme participants (trainees); participants' managers; and programme administrators. The order of importance (in terms of Guba and Lincoln) was not distinguished, however the technical vice president and senior manager group were treated as such.

3.3.2.5 Questionnaire

Easterby-Smith (1994) describes the questionnaire as 'a whole methodology' in that it is a class of methods rather than any single method. Questionnaires provide a relatively low cost method for collecting data from a large number of people. They can be applied for the measurement of knowledge and attitudes and used to collect data from which behaviour can be inferred. Questionnaire design and, in particular, question wording are crucial to maximising the validity of data obtained by a question asking process (Sudman and Bradburn, 1982).

Questions can be written in a variety of ways, however the main distinction is that of open and closed questions. Closed questions are written to limit the number of possible answers thereby facilitating data analysis. Open questions direct the respondent in the general area in interest, without prescribing the nature of the answers.

For this study, three types of questionnaire were used; to obtain pilot participant views on programme design; to measure changes in participants' knowledge; and to measure changes in participants' attitudes. Participants' views on the programme design were largely obtained using rating scales. For changes in participants' knowledge, a specific type of written test questionnaire was used and for measuring changes in attitudes, attitude scales were employed.

3.3.2.6 Rating Scales

Easterby-Smith (1994) distinguishes rating scales from attitude scales by the level of complexity of construction and interpretation; rating scales are simple and often employ one question item for each object, whereas attitudes scales are more complex. Attitudes scales are reviewed in section 3.3.2.8.

Basarab and Root (1994) identify rating scales as a useful means of collecting participant responses in a range of situations for a variety of purposes and distinguish four types of scales;

frequency; satisfaction; quality; agreement; and semantic differential scales. Frequency scales require respondents to indicate the how often an incidence of interest occurs. The question item is constructed with a simple scale continuum, typically ranging from 'all the time' to 'never' with descriptors in between (i.e. 'about 50% of the time'). Satisfaction scales are similar in construction, ranging from 'completely satisfied' to 'completely dissatisfied', as are quality and agreement scales. Semantic differential scales are constructed with statement items about an aspect of the training and respondents are required to indicate their view on a series of scale continuums from adjectives to their antonyms (e.g. fast to slow, with respect to pace of training).

Simple rating scales are pragmatic and easy to design and administer. They fulfil a purpose of general ordering of participant reactions and views with respect to training but have limited reliability and validity. Dane (1990) identifies three general issues for consideration when using rating scales; face validity; instructions; and item bias.

Face validity of scale items is concerned with the contents of the items and the extent to which they relate to the phenomena of interest. Oppenheim (1992) identifies many dangers with face validity (i.e. item representation of the phenomenon; balance of items with respect to content; and purity of items), however Dane emphasises the pragmatism of simple rating scales and suggests face validity is determined by the judgement of those who design the measure, and offers a general rule of thumb analogous to inviting people to a family reunion; 'if it ain't related, it ain't included' (p.264).

Instructions follow similar rules to those for questionnaires and are concerned with vocabulary, clarity and use of examples, however Dane expands on these by including the element of pre-requisite knowledge. Respondents should not be asked their views about issues of which they are unlikely to have the required knowledge (i.e. whether the national debt is too large requires knowledge of the national debt and gross domestic product - if a respondent does not know these, they can only guess).

Item bias refers to the extent to which the wording or placement of an item affect's someone's response. Dane suggests items should not be; 'double-barrelled' in that a single item contains two or more questions or statements; contain 'emotional flags' which reinforce a given position; or ambiguous by using relative terms such as big or small. With regard to placement, items are part of a series within a questionnaire and can contain implicit links with one another.

The rating scales employed as part of the pilot evaluation are very simplistic and as such are used for inferring general trends in terms of the identified aspects of the programme of interest to the development team. Whilst these were not explicitly selected as critical features of a training programme within the organisation, they were identified by experienced members of the development team and as such were generally considered to be those aspects which are important for ensuring a programme will be adopted by the organisation.

Category rating scales (Kerlinger, 1986) are also employed as part of the participant behaviour observation. The scales provide an observation framework which allows observers to record behaviour against pre-determined descriptors of interest. These scales were developed using critical incident technique (section 3.3.2.2).

3.3.2.7 Knowledge Gain Test

Knowledge tests are advocated by Kirkpatrick (1959b; 1994) as part of any level 2: learning evaluation. They are also very commonly used to make before / after course comparisons to evaluate training (Phillips 1991). Bramley (1991) offers a three level framework for the development of tests to measure job knowledge; (i) the basic level of isolated pieces of information; (ii) the arrangement of pieces of information into procedures; and (iii) the knowledge necessary to know when to apply procedures. Bramley's framework reflects the cognitive ordering of Bloom (1956) in that the presentation of knowledge progresses from simple abstract pieces of information to a more complex integration. Bramley distinguishes five types of knowledge tests; open ended questioning; short answer items; objective test items; multiple choice questions; and true false questions for use in training situations.

Knowledge tests are used as part of the implementation evaluation of the programme to measure changes in levels of knowledge as a product of training. A short multiple choice question format is used. Questions are determined against specified course objectives.

3.3.2.8 Attitude Scales

Oppenheim (1992) is widely referenced in the literature and provides a comprehensive analysis of attitudes and their measurement. He describes attitude as 'a (person's)[3] state of readiness, a tendency to respond in a certain manner when confronted with certain stimuli' (p. 174) which are; reinforced by beliefs; attract strong feelings; and may lead to particular behavioural intents. He elaborates an attitude in terms of its content and intensity, where content refers to what an attitude is about (i.e. race; war; and religion) and intensity refers to the endurance, depth and stability of an attitude. Oppenheim notes that social psychologists have generally distinguished the intensity of attitudes in terms of 'opinions', 'attitudes', 'values', 'basic attitudes', and 'personality'.

[3] In brackets added.

Oppenheim identifies four types of attitude scales (Bogardus; Thurstone; Likert; and Guttman) observing that they generally consist of six to twenty-four (or more) items which are usually attitude statements with which the respondent is asked to agree or disagree. The Bogardus social-distance scale (Bogardus, 1933) was developed to measure attitudes to ethnic prejudice using arbitrary scale items for ordering people's attitudes on a continuum. Thurstone scales (Thurstone and Chave, 1929) are attitude statements on a scale continuum derived from a panel of judges (up to three hundred people). Respondents are asked to agree or disagree with attitude statements as a measure of their attitude within the scale. Likert scales (Likert, 1932) consist of attitude statements against which respondents are asked to place themselves on a continuum (1 to 5: disagree to agree) for each statement and a measure of their attitude is derived from the set of items. The Guttman or scalogram method (Guttman, 1950) consist of series of progressive attitude statements on a scale. Respondents are asked to indicate those statements they accept and the ultimate accepted statement is used to derive a measure of their attitude.

For the empirical study, a Likert-type scale is employed to measure changes in trainees' attitudes with respect to quality and to the organisation. Likert scales are the most popular scaling procedure (Kerlinger, 1986) and are less laborious than Thurstone scales, but have been found to correlate well with Thurstone scales in experimental studies (Oppenheim, 1992).

Likert scales are developed from a composed pool of attitude statement items which are neither extreme or neutral in terms of the attitude under investigation. For each item a five point attitude continuum running from strongly agree to strongly disagree is used to collect participants responses. In the case of the scales used in this study, a nine point scale was adopted.

The points on the continuum are assigned 1 to 9, relating to the favourability of the attitude. Reliability of each item is established in terms of the overall total. Whilst Oppenheim notes that the item analysis should be undertaken by correlating each item with some reliable external criterion of attitude that the scale is intended to measure, this is almost never available. Therefore the total item pool is assumed to be the best available measure of the attitude. With this assumption, an internal-consistency measure (correlation coefficient) of the fit of individual items to the overall total item pool (minus the item of interest) and the items with the highest correlation are retained. Pearson's product-moment coefficient of correlation (Pearson and Hartley, 1954) is the most common measure of correlation coefficient (Harper, 1983) and was used primarily in this study.

3.3.3 Data Analysis Methods

Distinguishing data analysis from its methods of collection is largely artificial as the methods of analysis are determined by and determine the nature of the data collection method. For the purposes of this review I have included analysis as part of the review of the collection method, however one recurring exception to this in the literature is that for qualitative data. As part of the empirical study, stakeholder interviews are conducted to elicit programme effectiveness data from stakeholders. These interviews are tape-recorded and transcribed, producing reams of dialogue containing information of interest to the evaluation.

Patton (1990) describes qualitative data analysis as 'the challenge of making sense of massive amounts of data, reduce the volume of information, identify significant patterns, and construct a framework for communicating the essence of what the data reveal' (p. 372). He further observes that there are no absolute rules for qualitative analysis but guidelines and procedural suggestions which ultimately depend on the intellect and style of the researcher. Reliability and validity cannot therefore be determined using standard tests; each qualititaive study is unique and so 'analysts have an obligation to monitor and report their own analytical procedures and processes as fully and truthfully as possible'.

Patton identifies three procedures for analysing qualitative data; content analysis; case analysis; and inductive analysis. Content analysis involves identifying, coding and categorising the data to facilitate the search for patterns and themes. Case study analysis involves organising data by specific cases (i.e. individuals; programmes; institutions; or groups) for in-depth study, where the purpose is to gather information about each case of interest. Inductive analysis is drawing patterns, themes and categories from the data, and unlike case analysis where categories of interest are predetermined, they emerge out of the data.

For the programme evaluation, content and inductive analyses were conducted.

3.3.3.1 Content Analysis

Holsti (1969) defines content analysis as 'any technique for making inferences by objectively and systematically identifying specified characteristics of messages' and Krippendorff (1980) elaborates on Holsti's definition, making explicit the importance of context when drawing inferences from the analysis; 'content analysis is a research technique for making replaceable and valid inferences from data to their context'. Krippenhoff's definition also includes the notions of replaceability (reliability) and validity, however it is widely accepted in the literature that any analysis of qualitative data can ever be error free.

For the purposes of the training evaluation in this study, an initial coding frame (code-book) was developed from six pilot interviews. The interview transcripts were content analysed by the initial interviewer and themes were drawn from the data. Reliability was addressed by two other people using the initial coding frame to check its utility against a random sample of data. At this stage categories were added or deleted as necessary. Once the final coding frame was established, the author used it to code the whole data set. Two new judges were then trained in the use of the coding frame. The whole data set was then randomly divided in two. Each judge then coded each half and the percentage agreements between judges and to the norm (as established by the author) were calculated. All anomalies were discussed and a final agreement reached. This procedure is termed 'accuracy reliability' and it is suggested that it is the strongest type of reliability (Kripendorff, 1980).

The validity of the coding frame was assessed by a panel of programme trainers who were asked to code the interview data into 'categories of meaning', in the same way as the two interviewers had done. As the trainer panel had no prior knowledge of the code-book that had previously been developed, any differences between the coding frames were identified and resolved.

3.3.3.2 Inductive Analysis

For the study, an inductive analysis procedure was developed to identify the causal patterns within the transcripts. An agent/target organising framework was devised based on Stratton et al (1988): where agents are persons, groups or entities which are instrumental in causing change or bringing about an outcome; and where targets are persons, groups or entities which are influenced by the agent.

To determine the nature of the causal patterns, causal attributions were identified and extracted from interview transcripts using the definition adopted by Joseph et al (1993) as those 'statements identifying a factor or factors that contributed to a given outcome' and where 'a stated or implied causal relationship has to be present'.

Attributions were coded in terms of 'positive-negative' and 'actual-potential' dimensions. The 'positive-negative' dimension coded attributions according to whether they were referred to as positive or negative events. The 'actual-potential' dimension was used to code attributions which referred events which had happened or were about to happen or which the respondent anticipated might occur in the future.

The development and application of the analysis techniques is given in chapter 5, and the results in chapter 6.

3.4 Ethical Considerations

The purpose of data collection and its subsequent analysis is to gain a true understanding of phenomena, that is to say one which accurately reflects reality. This relies on the accuracy of the data collected which is dependent on the honesty of responses given, particularly in the areas of interviews, questionnaires, and observation.

If accuracy of information was the singular objective of an evaluative investigator, then by collecting data without subjects (respondents) knowing or consenting would increase the likelihood of accurate data collection. Douglas (1976) argues that the only way to gather data of any validity is for the researcher to operate in a covert manner.

For researchers, including those involved in evaluation, however, ethics are an important consideration as they have moral, legal and professional implications (Burgess. 1989).

Cavan (1977) defines ethics as:

> *"a matter of principled sensitivity to the rights of others. Being ethical limits the choices we can make in the pursuit of truth. Ethics say that while truth is good. respect for human dignity is better, even if, in the extreme case, the respect of human nature leaves one ignorant of human nature."*

3.4.1 Ethical Dimensions

Ethical issues relate to many aspects, including accuracy, confidentiality, breadth of consultation, rights of consent and access, and continuity of purpose (Raffe et al, 1989). Morrison (1993) argues that "the right of the public or stakeholders to have access to evaluation data and the right of the individual to privacy" is a fundamental tension which is central to the ethical issues surrounding evaluation.

Lawler (1998), drawing on literature related to ethics in education and other professions, identifies a range of ethical principles including; (i) recognising rights and dignities of individuals; (ii) developing human potential; (ii) providing employers, clients and learners with the highest quality education, training and development; (iv) complying with laws and regulations; (v) maintaining confidentiality; (vi) improving public understanding of human resource development; (vii) fair and accurate representation of one's credentials; and (viii) contribution to the continuing growth of society. Lawler comments, however, that little attention is given to ethical considerations in the literature on training evaluation and that there is a lack of guidance specifically designed for corporate training evaluation.

Easterby-Smith (1994) identifies three questions related to the ethics of evaluation; (1) what is the basis of the contract whereby one person is able to gather information from another?; (2) in what senses can such data be treated as confidential?; and (3) who does, or should own the results? which should be addressed for the evaluation of training.

Returning to Morrison (1993), ethics, like all other issues surrounding evaluation should be discussed and agreed before the evaluation begins.

From the literature and one's own experiences of training evaluation and other investigative activities in industry is that ethics are paramount to the perceived integrity of an individual or organisation. Apart from the problems of falling foul of governing authorities, the reputation of individuals and departments will be damaged (or enhanced) by the way in which it collects data, interprets and reports it. If one has to work in an organisation for 40 years and wishes to do one's job well, then the trust, respect and cooperation of work colleagues are overriding factors in the way evaluation is conducted.

3.5 Summary

Measurement is a system of using symbols which represent referent properties of an entity of interest in order to describe the entity and allow comparisons between entities. Referent properties are modelled into a coherent framework which is both valid and reliable. Validity is determined by the extent to which the model represents the referent properties and reliability is determined by the extent to which the model is capable of repeating representations of referent properties in a variety of conditions. In this sense, the principles of measurement are as applicable to the human or social sciences as they are to the natural sciences. The difficulty arises however from the complexity and our understanding of the entity under investigation.

Measurement plays a key role in evaluation as it enables us to distinguish important evaluand characteristics through the modelling process and provides a system for comparison of a characteristic of interest at various stages, indicating whether changes have occurred, and in some cases the magnitude of the change.

In the context of evaluation, methodologies for the collection and analysis of data are application of measurement system modelling. They provide the technology for evaluation and therefore are a limiting factor in the way we evaluate training. Evaluation methodology influences our thinking of how to evaluate, which in turn, influences our development of evaluation methodologies. Within the empirical study in this thesis, a range of evaluation methodologies are used to collect and analyse data.

Ethics provide codes of conduct both in a legal and moral respect. Within an organisation, the way in which evaluation is conducted is as important as the information it yields. Ethic conduct protects the reputation of an evaluation activity and ensures the rights of the individual are put before any other interest.

With respect to this study, ethical concerns were addressed since most of those involved in the evaluation are long term employees of the company and as such likely to be 'lifers'. Good reputations are developed through hard work and respect for others, whereas bad reputations are easily gained and difficult to lose.

Part II

The Empirical Study

Part II: The Empirical Study

'The knowledge of the world is only to be acquired in the world, and not in a closet'

- Lord Chesterfield, 1746

The empirical study was conducted over a five year period and concerns the design, development and implementation of an engineering and manufacturing quality training programme. The training programme (the evaluand) was implemented by a large multi-national manufacturing company and involved the training of approximately 6000 engineers in six European countries.

The goal of the programme was to change employees' knowledge, attitudes and skills in pursuit of the company's quality mission. The programme involved training in a comprehensive range of quality management and improvement strategies. Necessarily it was structured into a series training modules, or courses, which were interrelated as part of the overall programme curriculum. The programme was multi-faceted and complex by the nature of its content, development, implementation, and target audience; with each facet offering opportunities for evaluation.

Chapter 4 describes the programme's wide ranging content and the manner of its design, development and implementation. This provides an insight, from the perspective of an engineer-training practitioner, of the nature of the training programme in its industrial context.

The evaluation was conducted in 2 phases; pilot evaluation; and implementation evaluation. Chapter 5 describes the evaluation framework employed and provides a descriptive account of the development and application of the components of the evaluation. As part of the development of each component, validity and reliability issues are considered. A broader analysis of the overall evaluation is the subject of Part III of this study.

Chapter 6 presents the results of the evaluation and the conclusions drawn with respect to the training programme. The use of evaluation results are considered in terms of the overall management of the programme, drawing on the researchers observations as a participant researcher. Again, analysis of the evaluation in these terms is considered in Part III.

Chapter 4

Engineers' Quality Improvement Training Programme

4. Engineers' Quality Improvement Training Programme

> *'Signum scientis est posse docere*[1]
>
> *- Auctoritates Aristotelis*

The purpose of this chapter is to provide the reader with a succinct summary of the engineers' quality improvement training programme in respect of its content, structure, target audience, and the processes by which it was developed, implemented and managed.

4.1 Programme Overview

The programme is a 36 day training curriculum split into 2 levels; Core and Specialist (Figure 4-I). The curriculum covers the contemporary (1989 - 1995) range of technical quality methods considered to be appropriate for the company's product development and manufacturing business and consistent with its quality philosophy as described in the Corporate Mission Statement (see Introduction to this study).

The curriculum is structured into seven training modules, with four of the modules at two levels; level I courses are intended for all engineers and level II courses are intended for those engineers whose work requires them to have specialist knowledge of the quality method. Table 4-I outlines the type and mix of engineers employed in Europe in the design, development and manufacture of the company's products.

[1] 'The touchstone of knowledge is the ability to teach'

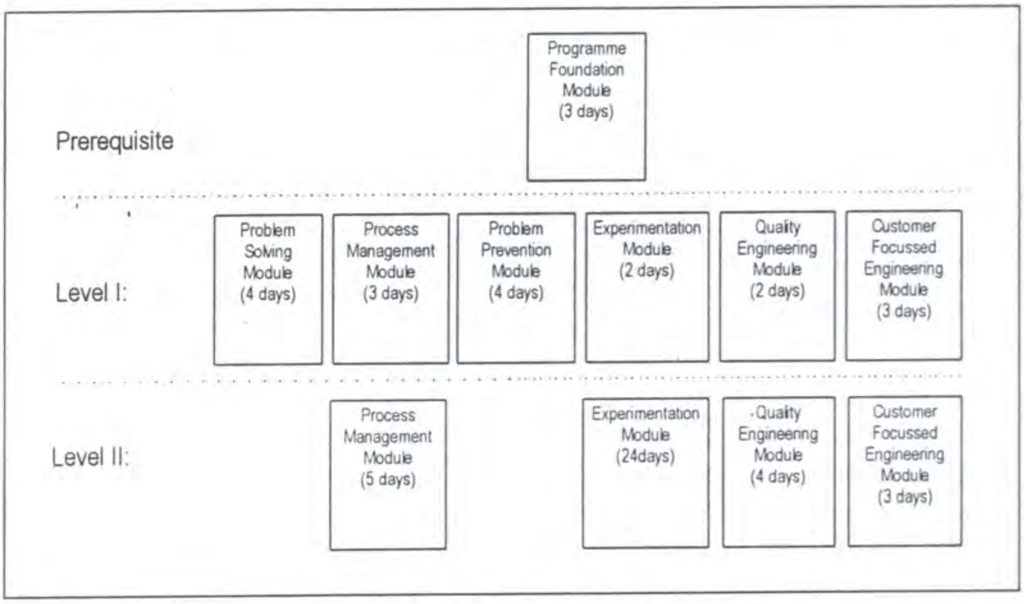

Figure 4-I: Programme Module Structure

Each module was designed as an integral part of the overall programme, something which had not previously been undertaken on this scale. Strong emphasis was placed on the linkages between the content of the modules which made up the programme.

All modules were designed in the context of the first module of the programme; Programme Foundation which provides the participant (or student) with an understanding of the underlying philosophy of the programme and a conceptual understanding of the quality methods and approaches and their relationship to each other and to the overall engineering process. This served to provide the participant with a framework or context with which to study the subsequent applicable modules of the programme.

	Design & Development	Manufacturing: Staff	Manufacturing: Plant
United Kingdom	1955	498	667
Germany	1431	327	581
Belgium	78	0	190
France	18	0	62
Spain	0	0	196
Portugal	0	0	106

Table 4-I: Type and Mix of Engineers Employed in Design, Development and Manufacture (1990)

To deliver the programme to the company's engineering community, primary programme delivery centres were established in Köln in (West) Germany and Boreham in the United Kingdom. In addition, secondary smaller delivery centres were also established in Setubal in Portugal, Bordeaux in France and Valencia in Spain. Each delivery centre was either located within the company's manufacturing facilities or close to them.

The programme delivery was undertaken by specially trained teams of trainer-consultants. These were engineers recruited from the company's product development and manufacturing operations divisions and trained specifically for the purposes of delivering the classroom based training aspects of the programme and providing guidance and advice to programme participants to facilitate post-training workplace application of the taught knowledge and skills.

Once selected, these in-house trainer-consultants were assigned to the Education and Training department for a nominal period of time and initially co-managed by the Product Development and Manufacturing Operations Training Managers. Following a restructuring of the management organisation in Europe, the trainer-consultants were subsequently managed by the newly appointed national training managers, who were responsible for both the Product Development and Manufacturing Operations.

On completion of their assignment to the programme, it was intended that the engineers would return to their mainstream functions, acting as programme experts to further facilitate the change process.

4.2 Programme Content

The programme content comprises a range of technical (quality) methods and tools' and behavioural (people) skills' intended to improve the engineering process. The primary technical methods are: disciplined problem solving, process management, problem prevention, experimentation, quality engineering, and customer focussed[2] engineering. The primary technical methods are supplemented by basic quality tools which also provide a foundation for many of the technical methods. The behavioural skills are; team building; communication; implementation; and innovation.

The following provides a brief introduction to the technical and people skills content of the programme. Programme content is contextually important to this study, however to avoid a lengthy account in the main body of this thesis, descriptive overviews of the content is given in appendix C, which summarises the training materials consisting of over 500,000 words, plus numerous graphics.

---

[2] deliberately mis-spelt

4.2.1 Technical (Quality) Methods and Tools

The following sections provide a brief overview of the technical (or so-called quality) methods and tools.

4.2.1.1 Basic Quality Tools

The basic quality tools are statistical and graphical tools for data collection and analysis. The set of eight tools comprise of what are commonly known as Ishikawa's (1982) Seven Tools; graphs, histograms (Guerry, 1833), cause and effect diagrams (Ishikawa, 1943), check sheets, Pareto Diagrams (Pareto, 1896), Control Charts (Shewhart, 1931) and scatter diagrams. In addition to these seven, flow diagrams were added by the programme development team.

4.2.1.2 Team Oriented Problem Solving

Team Oriented Problem Solving is a structured methodology for solving problems. Kepner and Tregoe (1981) define a problem as a deviation of a system's performance from its expected level. The Team Oriented Problem Solving Process was developed by the company as a standard corporate approach to problem solving, providing both a methodology for the resolution of problems and a common reporting format. The problem solving strategy comprises a rational process and process facilitation techniques.

4.2.1.3 Process Management

Process Management is a methodology for controlling and improving any process, although the emphasis in the programme is on manufacturing processes. Process Management is based on Deming's concept of 'Profound Knowledge' (Deming, 1993) where, for the continuous improvement of a system, two elements are required; an appreciation for the system, and knowledge of variation.

This concept is applied in Process management through four steps; a) identify and define the process; b) establish process management responsibilities; c) define and establish process controls; and d) improve process performance.

In the context of automotive manufacture, the finished product consists of over 500,000 components. Each has to be manufactured and assembled and each has characteristics which are important to the functioning of the vehicle. Controlling the processes by which the components are manufactured and assembled is essential to the overall quality of the product.

4.2.1.4 Failure Mode and Effects Analysis

Failure Mode and Effects Analysis (FMEA) is a method for anticipating potential problems, prioritising them, and preventing them or reducing the severity of their effects to reduce or prevent customer dissatisfaction.

The origins of FMEA lay in the aerospace industry (Samaio, 1995). Developed by NASA in the 1960's, FMEA was used to improve the reliability of aerospace, military, nuclear and electronic industry equipment and processes. Since the early 1970s, it has been used in the automotive industry as part of the product design and manufacturing planning disciplines, and it has been increasingly focused upon in the last 10 years by Western automotive manufacturers as a way of improving quality and reliability and reducing cost to compete with competition from Japanese manufacturers (Dale and Shaw, 1989).

The current methodology used in the automotive industry (SMMT 1989; SAE, 1994) is based largely on the American Military Standard MIL-STD-1629A (USA Department of Defence, 1980), although the company, as do other manufacturers (Aldridge and Dale, 1994), work to their own derivative of the methodology which forms part of the company's overall quality standard.

The FMEA methodology is applied during the product design or manufacturing planning stages of the engineering process. Concept or System FMEA studies and Design FMEA studies are employed to anticipate potential product design problems, and Process FMEA studies are employed to anticipate potential manufacturing process problems. FMEA is a time consuming task which, potentially, lends itself to automation (Price et al, 1992). At the time of the conception of the programme, FMEA remained to be a manual process.

4.2.1.5 Experimentation.

In the context of the programme, experimentation provides for the engineer to gain knowledge about a particular product or process. According to Groves and Davies (1992), statistically designed experimentation is a methodology whereby many design changes can be made at once and conducting a series of tests and evaluations before decisions are made as to what next steps are taken in the development of the product or process.

A product (or process) can be modelled in terms of its function (or outputs) and the factors which affect it. By understanding the complex relationships which exist between factors and how they effect function (or output), engineers can exploit these relationships to improve product quality. The experimental process is described in terms of the Deming cycle or Deming wheel (Scherkenbach, 1988) of plan-do-study-act.

4.2.1.6 Quality Engineering

Quality Engineering is an embracing term which refers to an alternative approach to engineering. It represents a significant paradigm shift from the traditional approach to engineering. Conceived by Genichi Taguchi (Taguchi, 1986), the engineering philosophy is to 'engineer in function, as opposed to engineering out problems' (Groves and Davies, 1992). Taguchi advocates the integration of five key concepts into the engineering process; a) energy transfer; b) ideal function; c) signal to noise ratio; d) robustness; and e) quality loss function.

4.2.1.7 Quality Function Deployment

Quality Function Deployment (QFD) (Akao, 1990) is a process for translating customer identified wants in a product into engineering specifications and subsequently into the product itself. Within the programme, QFD is presented within a wider framework termed Customer Focussed Engineering - the name which, perhaps, best describes the philosophy of this approach.

The definition and description of quality is explored further, with the introduction of Basic, Performance and Excitement Quality definitions. These distinctions between the various types of quality are necessary because of the underlying assumptions of the company's definition of quality; '"Products and services which meet the needs and expectations of customers".

4.2.2 Behavioural (People) Skills

The second primary topic area of the programme is the behavioural, or so called 'people' skills. Within the company, behavioural training is distinguished from technical training, both by the organisation of the Education and Training function, and subsequently, by the arrangement of these skill sets in the training catalogues. Taking a learner centred view-point, the engineer is concerned with skills which are employed in engineering. These not only include technical skills - the primary topic of the engineer's pre-employment education, but behavioural or people skills.

In 1991, the company employed over 100,000 people in Europe in the design, development, manufacture and sale of vehicles, including necessary support functions such as Personnel and Computer Systems. Engineering, therefore, is not an individual effort. Current day literature on Quality (Oakland, 1993) emphasises the importance of people and teamwork in the effort to improve quality. When the program was conceived however, this link, although in the minds of many practitioners was not an explicit feature of the literature.

Because of the reputation of 'behavioural skills' training and the negative perception held by many of the company's technical community (Brittle, 1991), the term 'people skills' was adopted to describe this topic area of the program. The development team's view was that behavioural were not regarded as important by the technical community, when compared to the technical skills. It was the belief of the team that the people skills were an essential enabler for the application of the technical skills. From the descriptions given in the previous pages the reader should have noted of the reliance of these techniques / methodologies of the input of several people and the importance of their knowledge of customers, products and processes to the quality effort.

For this reason it was decided that the people skills curriculum would be integrated into the technical skills curriculum. Later, in the account of the development of the programme, the reader will be able to better appreciate how this was achieved and the design advantages and drawbacks of this approach as perceived by the development team.

People Skills of the programme were eventually categorised into four topic areas; team building; communication; innovation and creativity; and implementation.

4.2.2.1 Team Building

The interpretation of team building was taken from a model developed by John Syer and Christopher Connolly (Syer and Connolly, 1987). Syer and Connolly's extensive background of working with sports teams to improve their performance through mental training influenced not only the terminology used to describe the skills, but the way in which they were trained. The team-building content of the people skills were eventually organised into six main topic areas; the effective team; roles and responsibilities; team process; the robust team; and right relationships

.

4.2.2.2 Communication

The second major topic area of the People skills; communication, is structured into 6 sections; listening, questioning, descriptive feedback, speaking guidelines, and framing information.

The emphasis is on verbal communication and is primarily concerned with the transfer of information and the common interpretation of that information. Each section encapsulates the concepts into techniques which are useable by the participants of the program.

Listening, or active listening as it is more commonly known, is included in the programme to address two main concerns identified in the communication which takes place at business meetings. The first is that listeners fail to understand the message that has been relayed to them through verbal communication. The second is that listeners ignore the message offering an alternative idea or statement.

The second section of communication is questioning which is drawn largely from Hargie (1988) and is divided in two main areas; questioning for content; and questioning for understanding.

The third area of communication; descriptive feedback refers to giving and receiving information about the performance of an individual or a group. The emphasis is on description, as opposed to judgement, as a method of giving feedback (Cairns, 1989). By stating observations of behaviour, the person receiving feedback is able to understand what his/her behaviour was, how it affected others, and give them something tangible to act upon. The receiver of feedback is also less likely to disagree with observations.

The fourth area of communication; speaking guidelines is intended to facilitate the communication process between team members. Based on Hargie (1991) and intended to encourage engineers to take ownership of ideas and criticism, a simple framework of nine guidelines is provided; talk from personal experience, speak to not about people present, address the person by name, look at the person you're talking to, say 'I', not 'we', make statements before questions, trace opinions back to observations, describe don't judge, accept that feelings will contribute to the discussion, and say 'I would', rather than 'you should'.

4.2.2.3 Implementation

The third main topic of people skills is implementation and is concerned primarily with bringing about change. The topic is structured into; force field analysis; action planning; decision-making; and change agency.

Force field analysis is set in the context of systems thinking (Senge, 1990) with the emphasis on viewing the overall process as opposed to the actual content. Developed by Kurt Lewin in the 1930's from his field theory (Checkland, 1981), force field analysis models phenomena as a system with forces acting upon it. Whilst all the forces are in equilibrium, the system remains stable and does not change, but when one force increases or another decreases, then change in the system will result.

Action planning provides a methodology for achieving specific objectives through the implementation of clear decisions. Action planning is undertaken within a clear agenda where substantial issues require follow-up. For the purposes of the programme, a set of rules were devised (see appendix C).

The third area of the implementation topic recognises the difficulty of team decision making, where a range of views are held by team members. Five types of decision making processes are considered; a) unilateral; b) polling; c) prioritising; d) compromise; and e) consensus. These are outlined in appendix C.

The principle aim of the training programme is to bring about change by providing engineers with different ways of designing and developing products and processes to improve quality. In this sense, the programme is subject to factors which influence the diffusion of new ideas in any culture. Based on Rogers (1983), change agency identifies people affected by change as innovators, early adopters, early majority, late majority, and laggards.

4.2.2.4 Innovation

The fourth topic area of the people skills is concerned with innovation methodologies of; thinking; scientific methodology; creative thinking strategies; idea mapping; cognitive mapping; language mapping; conceptual block-busting; brainstorming; paradigm Shifts; and innovative product development. The programme content draws largely on Ornstein (1972), De Bono (1991), Koestler (1964), Arnheim, (1969), Buzan (1993), Kelly (1955), and Kuhn (1962). Again, the reader is referred to appendix C.

4.2.3 An organising framework for the methodologies

The content is intended as part of an overall programme comprising of interrelated methodologies which serve a common goal; to improve product quality and increase customer satisfaction. Many of the methods have previously existed and been the subject of training programmes as individual initiatives with occasional token reference to other quality (technical) and people skills.

The development core team recognised the need to provide an organising framework for; a) the development of the training programme; and b) to serve as a learning aid for participants in beginning to understand the complex relationships which exist between the methodologies, skills and attitudes which form the entire programme. The foundation concept of the programme is the notion of quality and it was therefore essential that a comprehensive definition of quality was shared between members of the development team and ultimately, with the participants of the programme.

4.2.3.1 Traditional definition of quality

The long established definition of quality was concerned with the manufacture of products to specified characteristics.

> *'Conformance to engineering requirements as described in drawings, specification, and related documents'*
>
> *- Internal Durability, Quality and Reliability document (1980).*

The emphasis was placed on quality methods which better enabled the company's manufacturing activity to control quality in these terms.

4.2.3.2 New definition of quality

Although the traditional definition was superseded by a definition of quality which was couched in terms of the customer, the engineering paradigm had not shifted prior to the programme to reflect this definition of quality in the way products and processes were engineered.

> *'Customers define quality, customers want products and services that throughout their life meet their needs and expectations, at a cost that represents value.'*
>
> *- Company Quality Policy Letter A - 5 (1984)*

To further define quality in the 1984 terms, the concepts of positive and negative quality and upstream and downstream quality effort were introduced and explored in the programme.

4.2.3.3 Positive and Negative Quality

The traditional definition of quality is an exponent of negative quality in that it is concerned with the elimination of things going wrong. Negative quality is improved by the elimination of manufacturing and product development errors. The product which performs as intended by the designer is regarded as having a high level of negative quality.

Positive quality, is concerned with how well a product conforms with the requirement of the customer. Positive quality can only be achieved through the voice of the customer. For this reason, positive quality can only be affected during the early concept stages of a product's life.

4.2.3.4 Upstream and Down Stream Quality Effort

Perhaps the most distinguishing feature of the new quality philosophy, is the shift from detection to prevention. Methods such as Quality Function Deployment, Quality Engineering and Failure Mode and Effects Analysis, are initiated early in the engineering process - at the planning stages of a new vehicle. Together, the aim of these methods is to identify the requirements of a quality vehicle and engineer it in a way which avoids the occurrence of problems. These methods are referred to as upstream quality methods as they are applied in the former stages of the engineering process.

Further, Quality Function Deployment and Quality Engineering are methods which affect the positive quality of the product. FMEA and the other downstream methods are concerned with improving a products negative quality.

Methods such as Process Management and Team Oriented Problem Solving are applied in the implementation stages of the engineering process - when vehicles are produced and used. These are referred to as downstream quality methods, as the appear in the later, or downstream, stages of the engineering process.

4.2.3.5 Integration of Technical and Behavioural Skills.

The application of the quality and behavioural skills, including the attitudes which develop through the understanding of their underlying philosophy are not applied in isolation to each other. They form an overall engineering strategy; an approach to engineering which best employs the creative nature of engineers to design develop and manufacture high quality products, as defined by the customer.

As previously explained, presentation of these concepts in training programmes had tended to be in isolation of each other, with separate courses for each of the topic areas which had previously been promoted within the company. The majority of the content of the programme was not new. The programme was intended to distinguish itself from previous training by the way it integrated the quality methodologies with each other and integrated the range of people skills in the quality methodology processes. To this end, conceptual models were devised to illustrate the linkages or relationships.

4.2.3.6 Chronological Overlap Model

Framed in terms of the macro Engineering process, the chronological overlap model displayed the application of the quality methods on a time line (Figure 4-II).

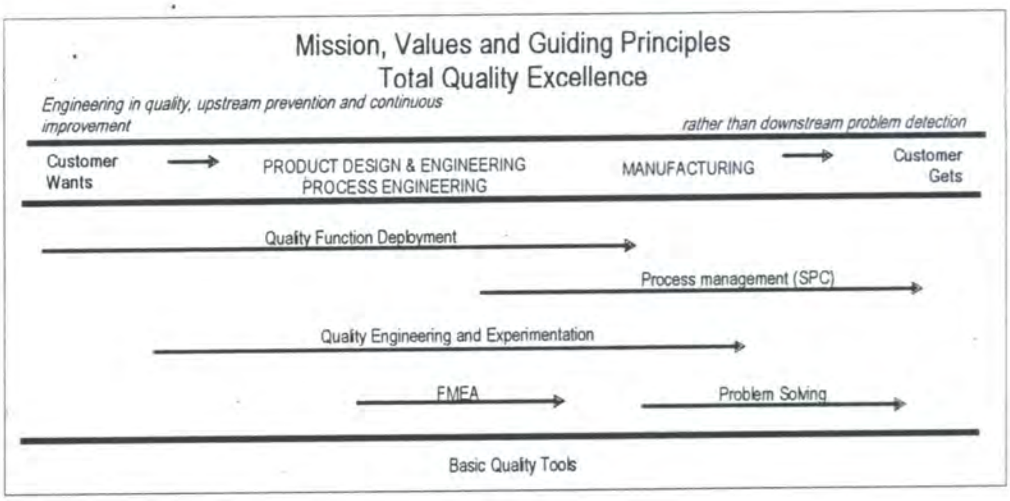

Figure 4-II: Chronological Overlap Model

At the top of the model the statements of 'mission, values and guiding principles' and 'total quality excellence' imply their umbrella context to the programme and its methods. The general product process is signified as in terms of customer quality as a process from 'customer wants' to 'customer gets'. Each of the primary technical methods are represented with arrows indicating their use within the general product process. Underlying the methods are the basic quality tools for data collection, analysis and communication.

Whilst the model provided a simple representation of the chronological order and overlap of the technical methods in terms of the general product process, it is incomplete in that it; (i) makes no reference to the engineering process; and (ii) does not represent the people skills. Despite these fundamental weaknesses, it portrays the overlapping nature of the quality methods and implies some relationship between them.

Its usefulness was such that it provided engineers with an introduction to the overall structure of the programme's technical methods in relation to the purpose general product process; to meet customer expectations.

4.2.3.7 Venn Diagram Model.

Henshall (1995) used a Venn diagram (Figure 4-III) to illustrate the overlap of the content area of the programme. By representing each of the quality methods using a circle and constructing the circles so that each overlapped every other quality method circle, Hensall implied there were interrelationships between all quality methods. He produced a similar diagram for the primary areas of people skills. Finally by overlaying people skills Venn diagram onto the quality methods Venn diagram, Henshall illustrated that their were interrelationships between the two sets.

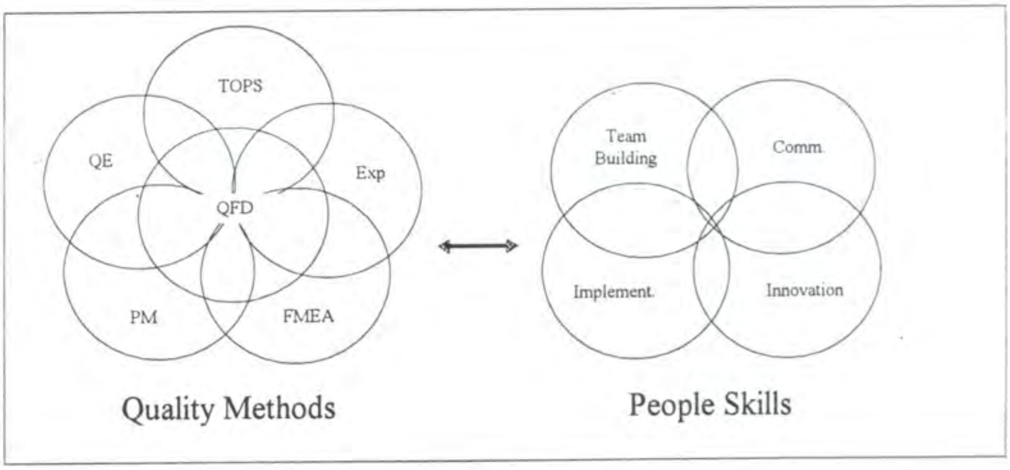

Figure 4-III: Venn Diagram Model

The multiple relationships, or linkages as they are referred to in the training material, are an important aspect of the programme, providing a systemic concept of quality improvement in engineering process. The move from stand alone quality methods and people skills represented a major shift in the approach to thinking about and providing training in quality improvement.

4.3 Curriculum Structure

The engineers' quality improvement training programme curriculum consists of technical (Quality) methods and Behavioural (People) skills. The curriculum is integrated, to reflect the integrated nature of the quality and people skills, but for convenience of delivery, it is modularised in to 7 training modules. A module is a training session of between 3 and 7 days duration, with 4 of the modules structured into two levels (Figure 4-I). The modules are structured in terms of the primary quality methods: Problem Solving, Process Management, Problem Prevention, Experimentation, Quality Engineering and Customer Focussed Engineering. As a pre-requisite to these a Programme Foundation module provides a conceptual overview of the programme philosophy and technical methodologies, together with some core people skills.

The people skills which are included in the programme are integrated with each other and are integrated into the technical skills (Table 4-II). The people skills concepts and techniques are structured into a 4 part structure, or curriculum; Team building, communication, implementation and innovation. The emphasis placed on the people skills is that they are stand alone, can be used in conjunction with each other and that they can be used as part of the application of the technical skills.

Within the context of the technical methodologies, the people skills are regarded as an enabler - allowing the application of the technical methods by a team to an engineering problem or opportunity. To reflect this intent, the people skills are structured into the programme around the technical methods, which form the primary focus of each of the 7 modules.

	Programme Modules (Technical Skills)										
People Skills	M1	M2	M3i	M3ii	M4	M5I	M5ii	M6i	M6ii	M7i	M7ii
Team Building											
The Effective Team	O										
Roles & Responsibilities	®	O									
Attitudes								O	O		
Team Process			O								
The Robust Team								O	O		
Right Relationships		O								®	
Communication											
Listening	O										
Questioning for Content		O									
Questioning for Understanding										O	
Descriptive Feedback		O									
Speaking Guidelines		O									
Framing Information					O						
Implementation											
Systems Thinking										O	O
Force Field Analysis					O					®	O
Action Planning	O		®							®	O
Decision Making Styles		O									
The Change Agent			O								
Advocating Change			O	O							
Innovation											
Thinking						O	O				
Scientific Methodology						O					
Creative Thinking Strategies						O	O				O
Idea Mapping								O			
Cognitive Mapping										O	O
Conceptual Block-busting							O	®	O		O
Brainstorming	O			O	®						O
Paradigm Shifts								O		O	O
Innovative Product Devt.											O

Key	
O	Module in which methodology is taught
®	Major reference included in this module

Table 4-II: The People Skills, where they are taught and referenced in the programme.

4.3.1 Programme Modules

The structure and duration of the programme is given in Table 4-III.

Module	Title	Level	Duration
Module 1	Programme Foundation	1 level	2.5 days (residential)
Module 2	Team Oriented Problem Solving (TOPS-8D)	1 level	4 days
Module 3	Process Management	Level I Level II	3 days 5 days
Module 4	Failure Mode and Effects Analysis (FMEA)	1 level	3 days
Module 5	Experimentation	Level I Level II	2 days 4 days
Module 6	Quality Engineering	Level I Level II	2 days 4 days
Module 7	Customer Focussed Engineering (CFE)	Level I OR Level II	3 days OR 4 days

Table 4-III: Curriculum Structure and Duration

4.3.1.1 Programme Foundation Module

The programme foundation module provides a conceptual level overview of the technical skills covered in the entire programme. In addition, the 8 basic quality tools are considered at an applications level. A selection of people skills are also included. The content of the module is as follows:

Conceptual overview of methodology and application:
The prime emphasis of the module is to provide participants with a conceptual awareness of the major quality methods in the programme; Team Oriented Problem Solving (8 Disciplines); Process Management; Failure Mode and Effects Analysis; Experimentation; Quality Engineering; and Customer Focussed Engineering (through Quality Function Deployment). The module also explores the conceptual linkages between these methods.

In addition, the seven Ishikawa tools and flowcharts are included at an applications level of training. This training is intended at a level where participants are expected to be able to apply the tools within their job immediately following training.

People skills:
At an application level of training the following people skills are the topic of the Foundation module; team building; team roles and responsibilities; Attitudes; warming up / warming down; holding effective team meetings; introduction to task, maintenance and process of teams; listening skills; introduction to brainstorming; action planning; and introduction to descriptive feedback.

4.3.1.2 Team Oriented Problem Solving (8 Disciplines) Module

The TOPS(8D) module is intended to provide participants with applications level training in the use of the TOPS methodology. The course is structured around an depth case study, with sessions where the concepts and methodology are taught and then applied by the class to the case study problem. Along side the 8 disciplines, the accompanying process helps are taught . The people skills elements are integrated into the problem solving process. This is intended to a) teach the people skills and b) link them to the technical skills. The people skills are presented as being generic and not just for use within a problem solving context.

Technical skills:
The technical skills taught in this module are; the eight disciplined approach to problem solving; decision making; concerns analysis; and introduction to problem prevention

People skills:
The people skills are; team roles and responsibilities; questioning for content; descriptive feedback; right relationships; speaking guidelines; decision making styles and practice in the application of task, maintenance and process in a team meeting context.

4.3.1.3 Process Management Module

The Process Management module is split into two levels and is intended to provide participants with applications level training in the use of Process Management methodologies. The level 1 course is a prerequisite to the level 2 course.

To assist in understanding, Process Management is divided into a 4 step approach of 1) Identify and Define the process, 2) establish process management responsibilities, 3) define and establish process controls, and 4) improve process performance. The content of the modules is as follows:

Technical skills:
At level I the conceptual framework of process management is developed from that previously considered in the Foundation module and the general principle of process management are taught in terms of this 4 step approach; identify and define the process; establish process management responsibilities; define and establish process controls; and improve process performance. In addition the level I module covers; the concept of statistical process control (SPC) and the relationship between process and capability; the concept and method for potential process capability (PIST - percentage of inspection points which satisfy tolerance and PIPC - percentage of inspection points which are process capable).

At level II, the range of tools and techniques available for process control are explored further. These are; chart control techniques for short runs, fixed tooling and multiple characteristics; CUSUM (cumulative sum) charting; and gauge and process capability.

People skills:
The people skills are common for level I and II and are; team process; team roles and responsibilities; use of descriptive feedback as the voice of team process; the change agent; and equal time for people.

4.3.1.4 Failure Mode and Effects Analysis Module

The FMEA module is intended to provide participants with applications level training in the use of three types of FMEA: Concept FMEA, Design FMEA and Process FMEA. To assist participant understanding, the FMEA methodology is divided into 6 stages; 1) Define scope and function, 2) identify potential failures (in function), 3) prioritise potential failures, 4)select and manage subsequent actions, 5) observe and learn, and 6) document the process.

Technical skills:
The technical skills content of the FMEA module are; defining function using function tree diagrams; using generic categories of product and process failure modes to identify potential problems; applying rating tables for probability of occurrence, likelihood of detection and severity of effects; and compiling FMEA reports using company software.

People skills:
The people skills content of the FMEA module are; brainstorming; force field analysis; and framing information.

4.3.1.5 Experimentation Module

The experimentation module is split into two levels and is intended to provide participants with applications level training in the use experimental methods.

Technical skills:
The technical skills content of the experimentation module I are; appreciation of the power of directed experimentation; demonstration of the advantages of multi-factor experimentation over 'change-one-thing-at-a-time' experimentation; design and apply experiments with factors at two levels; and management of the experimentation process.

At level II, the technical skills are; design and apply experiments with factors at 3 levels; the role of control factors and noise factors; to recognise the concept of robustness as an interaction between control and noise factors; and to utilise the concept of signal to noise ratio as a measure of robustness.

People skills:
The people skills for experimentation are common for levels I and II and cover; creativity and innovation; conceptual block-busting; and visualisation.

4.3.1.6 Quality Engineering Module

The Quality Engineering module is split into two levels and is intended to provide participants with applications level training in the use quality engineering principles and practices. The stated content of the modules is given below and is common for level I and II. The material states that the content is 'introduced at level I and then examined in more detail at level II'.

Technical skills:
The technical content of the module covers; the engineering method; new technology versus competitiveness technology; signal to noise ratio; the engineered system; the dynamic approach; measurement; translating problems into variability; the loss function; quality engineering methodology overview; parameter and tolerance design; response characteristics; ideal function; strategy for improvement; dynamic experimentation; and operating window.

People skills:
The people skills content of the module covers; mind mapping; visualisation; paradigm shifts; creativity and innovation; and emotional blocks

4.3.1.7 Customer Focussed Engineering Module

The Customer focussed engineering module is split into two levels and is intended to provide participants with applications level training in the use of the customer focussed methodology as part of the engineering process. The module is structured around the quality function deployment methodology and a substantial case study which provides participants with the opportunity to apply the methodology to a simulated engineering project.

Unlike the other 2 level modules of the programme, the level I is not a pre-requisite of level II in module 7. Participants attend either level I or level II depending on their role within the total engineering process.

Technical skills:
Significant emphasis is placed on the content of all the previous modules and how they fit together as part of the engineering process. At level I the technical skills content is; the context of quality function deployment; QFD methodology; and Fukahara's four phase approach to product engineering.

At level II the content is; pre-planning; Pugh concept selection; function trees; descriptive statistics; quality engineering; experimentation; FMEA; SPC; gauge studies; process capability; TOPS (8D); and the basic quality tools.

People skills:
The people skills are common for levels I and II and are; cognitive mapping; language mapping; questioning for meaning; force field analysis; right relationships; action planning; systems thinking; creativity; and innovative product development.

4.3.2 Teaching Case Study

The program employed a major case study, the Cyclone Fan Company, throughout each of the training modules. The aim of the case study was to support the integrative nature of the program by providing a common framework in which the programme's techniques would be used.

The stated objectives of the case study were to: i) Provide a common, unifying element in the programme, ii) Exemplify the concepts taught in the programme, iii) Demonstrate the practical application and relevance of the techniques taught in the programme, iv) Demonstrate the relationship between concept, framework and technique, v) Enable students to develop a competence in these techniques, vi) Give participants practice in addressing problems in a multi-functional team, vii) Enable participants to develop team working, interpersonal and presentational skills, and viii) Provide a 'neutral territory' that will nevertheless develop insights and competencies that can be directly related to the participants' job responsibilities.

The case study followed the fortunes of a group of executives and other employees of a multinational company as it moves from a defect detection operating quality philosophy to one of total quality excellence. The company and the characters portrayed demonstrate exaggerated (or in some cases not so exaggerated) attitudes and behaviours which mirror those found in the company. Through an external consultant, the company is transformed by adopting a new quality philosophy and the tools and techniques (technical and people skills) which are the subject of this training programme.

The case study is introduced as part of the first Programme Foundation module and then appears in each subsequent module providing scenarios in which participants in the programme address issues and apply the skills they are learning. The case study is very elaborate and uses text, videos and simulations.

4.4 Programme Development

The nature of the content of the program and the scope of its organisational impact, determined that its development required a cross-functional multi-national effort. The scope and complexity of the program also deemed it necessary that the development followed a predetermined process. The following outlines the composition and structure of the development team and the process by which the programme was developed.

4.4.1 Programme Development Team

The development team was established with representation from the education and training function and the Quality Office. The training representation comprised of representatives from the

Product Development and Manufacturing Operations European Education and Training Staffs and the UK and German national Training staffs, reflecting the then complex organisation in Europe. The Quality Office representation comprised of representatives from the PD and Manufacturing Statistical Methods Offices and the European Quality Office. The team also consisted of external subject matter experts in both the technical and behavioural topic areas and training materials production specialists.

4.4.1.1 Team Structure

The development team was organised into a small core (leadership) group, with the wider development team being structured into mini-teams, or module sub-teams (these terms were used interchangeably. Leadership for the mini-teams were provided by members of the core team.

4.4.1.1.1 Core Team

The composition of the core team changed during the development of the program (1989 - 1993). Initially core team membership consisted of seven members representing; Statistical Methods Office (Manufacturing Operations); Statistical Methods Office (Product Development Group); European Quality Office; UK National Training Staff; Germany National Training Staff; Manufacturing Operations (European) Training Staff; and Product Development Group (European) Training Staff.

Four of the seven members of the team were assigned almost full time to the development of the programme and were co-located. The remainder of the team worked on a part time basis through development meetings.

The initial work of the team focused on establishing a development process and identifying the skills, knowledge and experience required to bring the process to a successful conclusion. Much of the latter work focused on project management and trainer recruitment and training.

4.4.1.1.2 Development Mini-teams

Subject Matter Experts for the program were selected from the company's existing consultant base. The selection process was informal and relied on the opinions of members of the core team. These subject matter experts, who in the main were external to the company, formed the wider development team. Expertise was drawn from; Corporate Quality Office (Dearborn) for Quality Function Deployment; UK based external training consultants for behavioural (people) skills, Statistical Process Control, Statistical Methods, Failure Mode and Effects Analysis, and educational design and authoring; USA based external consultants for Quality Function Deployment and Taguchi Quality Engineering. In addition, UK-based consultants were employed for; language translation; copywriting and typesetting; and visual teaching aids design and production. Other in-company experts were brought into the team from time to time for; communication process design and facilitation; and Dimensional Control Planning (process control method).

In addition, the team was supported by staff members of the respective external companies, and so there were up to 30 people working on the development of the program at a time.

To manage the development, the team was organised into mini-teams. The organisation structure was based largely on the modular content structure of the programme. Certain aspects of the content and development made it necessary to have min-teams not directly aligned to the modules. These were for the people skills, case study and training materials.

Initially, the team structure (Figure 4-IV) was informal and agreed between the members of the whole team. This was subsequently formalised in 1991 as part of the major reorganisation of the company in Europe resulting from the 1990 Simultaneous Engineering Study.

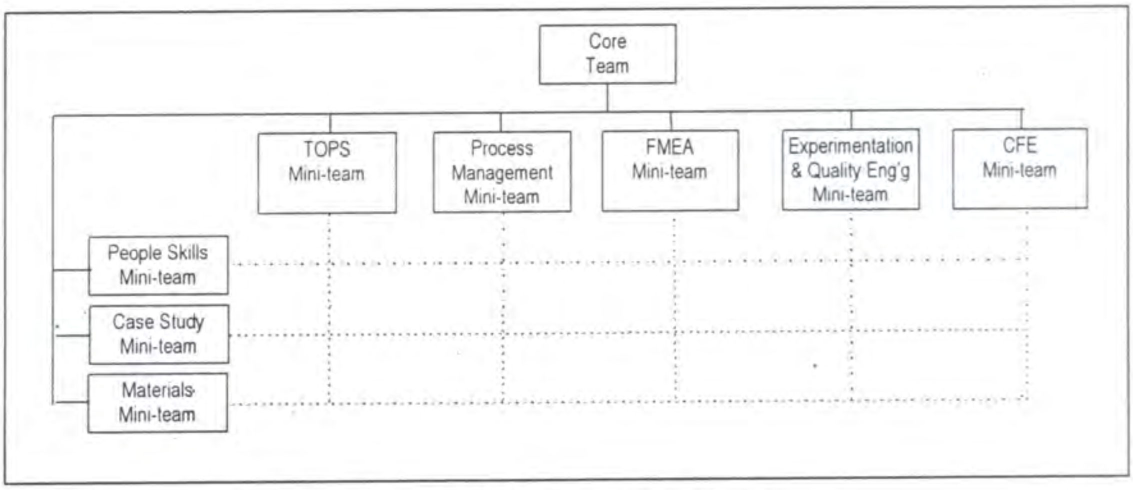

Figure 4-IV: Development Team Structure

4.4.2 Development Process

The development of the programme was launched when the first full team meeting was held on 14th and 15th February, 1990.

The development of the programme was intended to follow the macro process shown in Figure 4-V. The 12 stage process outlined the steps of development from the initial needs analysis through piloting and review to full scale launch. The process documented is taken from archives of the inaugural full-team meeting held on 14th and 15th February, 1990. Whilst the development broadly followed this process, the actual process used was very much more complex and interactive. Materials were restructured and redesigned, based primarily on input from trainers and design team members. In some cases, materials were re-written 11 times before they were considered to be acceptable. The process of developing training requires a good understanding of the content and as understanding improves, then better ways of expressing concepts become apparent.

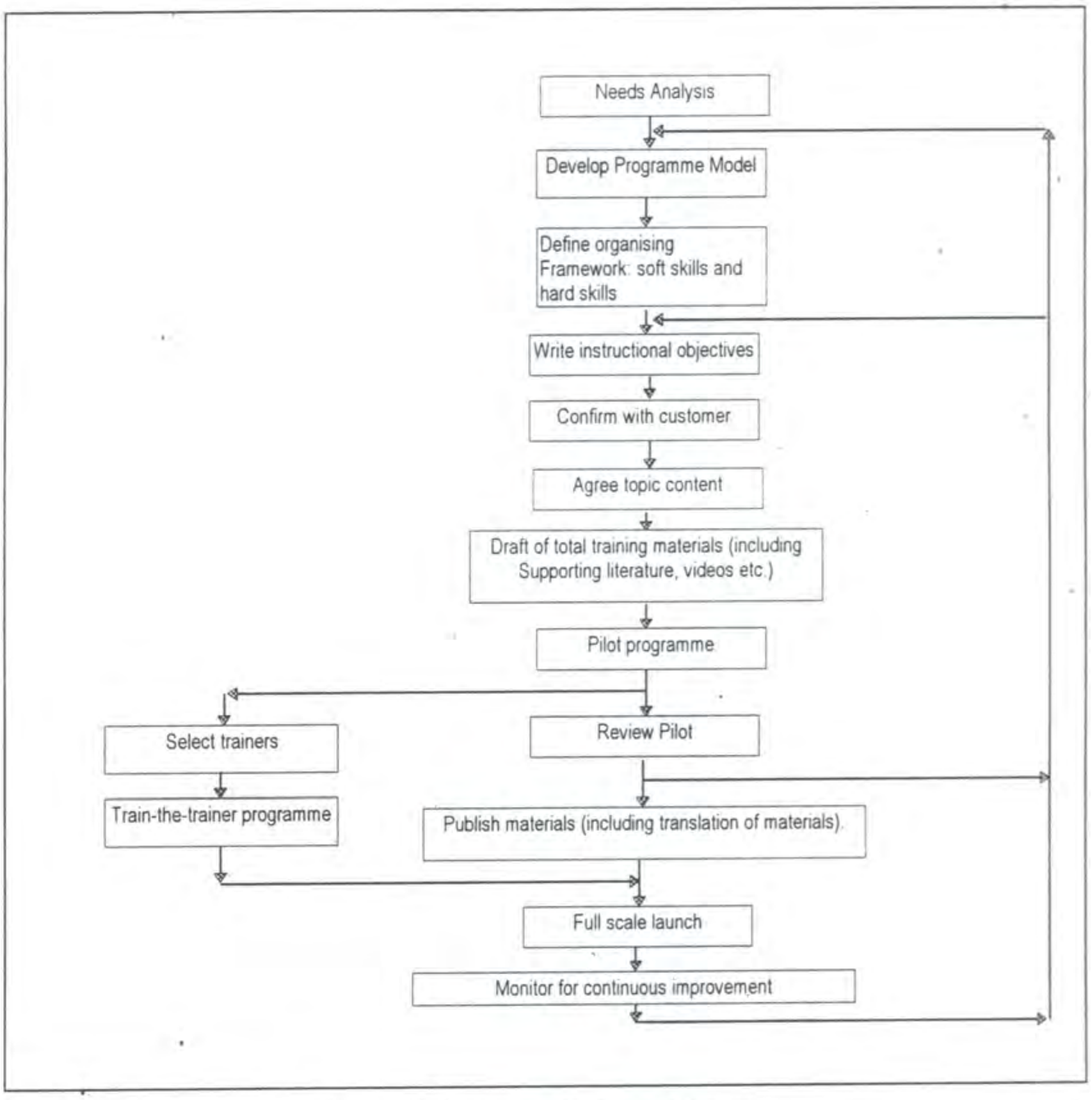

Figure 4-V: Programme Development Process

4.4.2.1 *Development by objectives*

During the initial stages of the development, it was agreed that the programme would be developed using behavioural learning objectives. Due largely to the experience in the team, Bloom's Taxonomy of behavioural objectives (Bloom, 1956; Krathwohl, 1964) was selected as a guiding framework. A description of Bloom's taxonomy, together with a review of the literature is given in Chapter 1.

It was intended that the objectives prepared for each of the modules would serve three main purposes: i) to provide a design specification for the instructional elements of each module, ii) to be used to compare and develop the content linkages between the modules, and iii) for the simultaneous writing of the case study. Whilst development of the behavioural objectives provided the general direction and shape for the design and development of material they were found to be time consuming in the short term. As management pressure to launch the programme increased, they became less of a priority in the minds of the course designers and were not kept up to date in line with changes to the programme (Brittle, 1991).

4.5 Programme Implementation

To deploy the programme to the company's 6000 engineers across Europe, considerable resources, facilities, and funding were required. Although accustomed with pan-European training initiatives, the scale of the programme would require the company put in place a management structure which would address the process, financial and political considerations.

Furthermore, by its innovative and complex nature, the programme would need to be delivered by specialist trainers who would not only play a role in the classroom, but serve as internal consultants who would provide workplace support to participants applying the methodologies to the engineering process. The trainers would also require programme specific training in the use of the training materials.

4.5.1 Programme Steering Committee

To manage the overall implementation of the programme, a steering committee was established under the direction of the company's European Education and Training Director. This group was an evolution from the 1988 Quality Education and Training Committee and was represented by the European Training managers and line managers from the engineering functions.

The role of the committee was twofold; 1) to define the delivery strategy and lever the necessary resources, 2) to manage the ongoing strategic issues of the programme.

The committee met monthly and the agenda of each meeting was determined by the issues which had arisen. The meeting also had standing items where programme status reports were given. These tended to focus of the numbers of engineers trained at the various delivery centres.

4.5.2 Training Delivery Strategy

To deliver the programme, programme training centres were established in the UK, Germany, France, Spain and Portugal. Each centre would have its own resource and be responsible for delivering the training to its national engineers. The delivery effort would be co-ordinated centrally through the European Education and Training staffs. All matters pertaining to ongoing development and the translation of materials would be managed by the programme core team.

Programme delivery was to be targeted at cross-functional program or engineering teams as an integral part of the development of new products.

4.5.3 In-house Trainer-Consultants

As the programme was specifically about engineering and would involve the facilitation of the application of the content of the programme to the engineering process, it was decided that the trainers would be recruited from the company's engineering functions. They would have experience of engineering and would not only understand the issues associated with the change, but would have the credibility to talk about the engineering process.

The engineers were initially to be seconded to the programme team for a period of 2 years. On completion of their secondment, they were to return to their engineering functions to resume an engineering position. It was expected that with their knowledge and experience of the programme, they would become a local source of expertise thereby continuing to affect the change process. Despite this intention, many of the engineers remained with the programme until they either retired, were promoted, or until the programme was replaced by the world-wide Technical Education Programme, where their services were no longer required in an ongoing training and consulting capacity.

This model of training delivery would require an extensive training programme for the engineers selected that would encompass not only the specialist knowledge required about the content of the programme, but also training and consulting skills.

4.5.3.1 Trainer-Consultant Selection

In-house trainer-consultants were selected using an assessment centre process. Assessment centres allow for the selection of candidates using an agreed selection criteria. Simulations or activities and interviews are used to collect data about each candidate which is assessed against the selection criteria.

Assessment centres were first used during WW2 to select intelligence agents. Since the 1950s, the assessment centre process has been used for selection increasingly, following AT&T applied it to select industrial managerial potential. In 1989, 58.9% of a sample of UK companies reported using assessment centres.

The selection criteria for the programme trainers was established using Critical Incident Analysis - a technique originally developed by Flanagan (1954) which is aimed at obtaining a record of specific behaviours from those in the best position to make the necessary observations and evaluations. The collection and tabulation of these observations made it possible to formulate the critical requirements of a trainer-consultant by identifying and analysing the critical incidents of an activity, namely training and consulting.

Using the core team, the critical Incident technique was used to identify the following requirements as being: Team Involvement; Resilience; Task Awareness; Presentation Skills; Interactive Skills; Process Awareness; and Commitment to the Programme. A description of each of these requirements, or characteristics, is given in Table 4-IV.

For each of the characteristics, or selection dimensions, positive and negative behavioural indicators were identified. These were descriptions of good or bad behaviour which would be used to aid observation and classification of candidate behaviour during the assessment centre activities.

Using the characteristics, assessment components were designed. These components were activities including making a short presentation, solving an engineering problem as part of a team, and writing a report. For each activity one or more of the dimensions listed could be assessed and each dimension was assessed more than once across the range of activities which formed the assessment centre.

The assessment centre was used to recruit trainers in the UK, Germany, France, Spain and Portugal (see Table 4-V).

Characteristic	Description
Team Involvement	Trainers work closely with all members of the team. They step in and help colleagues when they are having difficulties (even when its not their specialist area) and contribute to team discussions and the development of the team. They recognise their own strengths and weaknesses and do not train or consult in areas in which they are not competent. They accept responsibility for failures in programmes and do not off-load onto others. They are willing to change if necessary. |their behaviour is consistent and does not fluctuate and they always portray themselves as the type of people who are committed to the team.
Resilience	Trainers are single-minded and determined to succeed. They confront problems head on and handle difficult situations without aggression or without losing control. They do not become defensive when questioned or challenged. Their determination to succeed does not cloud judgement and they never compromise skills or knowledge when challenged. They work in a relaxed open style, never openly argue with participants however awkward and never forget they have responsibility to the group as a whole.
Task Awareness	Trainers have objectives to achieve, they work to a time table and understand the importance of logistics. During training or consulting projects they research the level of knowledge of the audience and regularly check for understanding. They are aware of the importance of making complex subjects easier to understand and know when to summarise and direct progress in order to keep people to task. They review progress and follow-up on matter arising. When questions are raised they rarely go off on tangents and always respond appropriately and with satisfactory answers.
Presentation Skills	Trrainers care about the appearance of the materials they use and illustrate their point with up-to-date diagrams, bullet points and models. They prepare for presentations thoroughly, speak in a clear voice and engage people they interact with by talking clearly and enthusiastically., relating input to the visual aids and notes. They set up problems for discussion with relevant examples and are able to highlight examples of behaviours in the group that illustrate learning points.
Interactive Skills.	Trainers like to let the group get on with the task and do not interfere other than to offer help or advice; they do not compromise themselves and are not direct with customers. They get the best out of people by listening, digesting and rationalising before responding or by referring questions back to them. They encourage participants to contribute by using positive language, inviting feedback, asking open questions and by reflecting back to the group so that they can build on issues. In turn they provide feedback in a descriptive, non-judgemental manner and are sensitive when dealing with delicate or difficult matters. Their interactive style is flexible enough to change according to the needs of the clients they are dealing with, treating managers and operators in the same positive way. They introduce themselves in new social situations and put people at ease by 'warming' them up and 'warming' them down by being relaxed.
Process Awareness	Trainers observe the behaviour within the groups and are able to evaluate when it gets in the way of effectiveness. By doing so they are able to pick up significant changes in the group's behaviour such as hidden agendas, changes in the group's feelings towards the training and when the group are off track and in need of assistance. They realise when they are communicating effectively with the group and when they are not. Above all, they are able to distinguish between normal learning problems and genuine concern and can identify where learning blocks have occurred and how to resolve them. Trainers are comfortable with silences and have a sense of timing enabling them to know when and when not to intervene in the process
Commitment to the Programme	Trainers know the fundamentals and have a good feel for all the training material. They are full of factual information and have a key to understanding of all the basics. They think conceptually, can quickly identify and solve problems, have a burning desire to be effective and look at the long term application of the programme. They believe in steering people to what they already know and seize opportunities for participants to learn about different concepts. They themselves are eager to learn. To this end they are keen to share successful and unsuccessful applications and frequently invite feedback from people they have helped or advised. The Trainer contracts with clients the boundaries of the projects and know when to hand completed projects over to the client. They do so in order that clients do not become over-dependent on them.

Table 4-IV: Trainer Selection Requirements

Country	Number of candidates assessed	Number of trainer-consults selected
UK	48	11
Germany	23	8
France	9	2
Spain	7	3
Portugal	12	4

Table 4-V: Trainer - Consultants

4.5.3.2 Trainer-Consultant Training: Level 1 modules

An extensive train-the-trainer programme was devised and implemented. For each module, 2 trainers were required to deliver the programme and so it was decided that each trainer would deliver the Foundation Module plus two other specialist modules. By consensus between the core team and the UK and Germany trainer-consultant groups, the specialist modules were paired. Together with the programme Foundation module, the pairings formed trainer sets of modules which are given in Table 4-VI.

Trainer Set	Programme Modules
A	Programme Foundation Process Management (Levels I & II) Failure Mode & Effects Analysis
B	Programme Foundation Experimentation (Levels I & II) Quality Engineering (Levels I & II)
C	Programme Foundation Team Oriented Problem Solving Process Management (Levels I & II)

Table 4-VI: Trainer sets of modules

The train-the-trainer programme comprised of teaching about the programme and its content, training on training and facilitation skills, instruction on how each module should be delivered, co-delivering the module with a members of the core team and finally delivering the module under the supervision of members of the development team leading to accreditation to deliver the module.

The duration of the complete train-the-trainer programme for each trainer was 6 - 9 months depending on the scheduling of the trainers sets of modules. The flowchart in Figure 4-VI provides an overview of the train-the-trainer process.

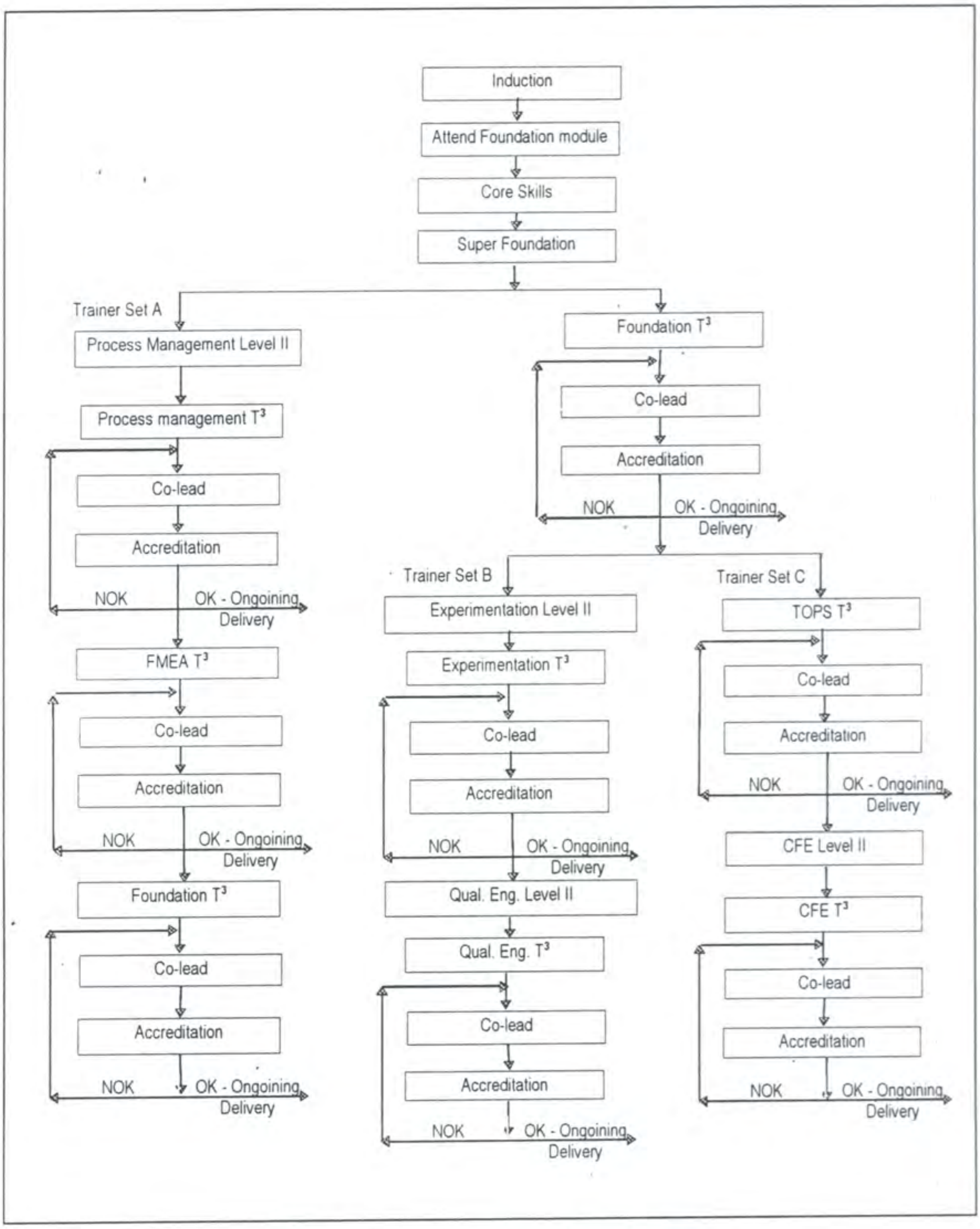

Figure 4-VI: Trainer-Consultant Training Programme

The trainer-consultant process followed a nine stage process which was initiated on joining the programme team and continued through their secondment period. By the nature of the training, trainer-consultants started to teach the programme and provide work place advice to trainees before they completed all stages of the training process.

Induction
All selected trainers underwent a 2 day induction to the programme. The primary goal of this session was to introduce trainers to the programme, to each other and to the development team.

Attend Foundation Module
As an extension to the induction, the trainers participated in the Programme Foundation module as participants. The main goal of this session was for the trainers to gain a conceptual understanding of the programme - just as for any other participant.

Core Skills
All trainers attended the 2 day core skills session. The main goals of this session were to provide them with training and facilitation skills.

Super Foundation
The super Foundation comprised of all of the level I modules of the programme delivered back to back. Again, the trainers attended this as participants with the goal of learning about the content of the programme. The Super Foundation was delivered over a 6 week period.

On completion of the super Foundation training, the trainers were divided into their Trainer sets A, B, and C. Set A would then commence their specialist module training before undertaking their Foundation module training. Sets B and C would undertake their Programme Foundation training before commencing their specialist module training. This sequencing was important as training needed to be made available to the engineering community as soon as possible. As the Foundation module was the first, followed by TOPS, these two were selected first for the trainer-consultant training.

For all modules, an identical process was used:

Module Train-the-trainer (T^3)
Also referred to as the master class, the T^3 provided instructional training for the trainers. The module was broken into its component parts as considered in detail. The duration of the T^3 was typically twice that of the module. Instructor guides were developed for and used as part of the T^3.

Co-Lead
Each of the trainers co-delivered the module with members of the development team to participating engineers. Their delivery was observed and the observations recorded. At the end of each day and on completion of the module, the trainers received descriptive feedback on their performance.

Accreditation

Similar to the co-lead, the trainers delivered the module to actual engineers participating in the programme, however they delivered the entire module between the team of two trainers without the assistance of members of the development team. Their delivery was observed, recorded and assessed against the criteria used for the assessment centre. Observation and assessment was undertaken by members of the development team. On completion of the module, a half day feedback session with conducted with each trainer. Where a trainer was considered to be proficient in delivering the module, they were accredited and commenced delivering the module unaided. The decision as to whether a trainer was proficient was subjective, but based on the observations and discussions with the trainer.

If a trainer was not considered to be proficient, then he / she would undergo further co-lead sessions until they felt the were ready for a second accreditation delivery. Where this occurred, individual coaching was given.

Once a trainer was accredited in a module, they began their training for the next module, as well as delivering those module(s) for which they had been accredited.

Trainer-Consultant Training: Level II modules

Due to a) the level of demand for level II modules, and b) the necessity for the trainers to meet the demand for the level I modules, these were initially delivered to the engineering community by members of the development team. When, eventually, the trainer-consultants were trained to deliver level II modules, a similar process to that employed for level I training was adopted.

Trainer-Consultant Training: Consulting

As part of the original concept for the implementation of the programme, the trainers were to play a dual training / consulting role. This consulting role had been regarded by the core team to be an essential element in transferring the skills taught in the programme to the engineering process. In reality, this vision was not achieved to the extent intended; the pressure to deliver training and meet audience targets which were the interest of the company's management gave trainers little time to spend in the workplace. This situation was made worse by the reluctance on the part of some trainers to act as internal consultants.

A two day consulting skills workshop was, however, used to train the trainers in consulting skills using a simple process model, where a contract was established between the trainer-consultant and the customer engineering area. It was intended that this would serve to ensure the integrity of the consultant's role was maintained and provide a record of achievement. Despite efforts by the core team, little evidence is available of consulting undertaken by the trainer-consultants.

4.6 Programme Management

The programme was managed by a programme steering committee comprising of Engineering and Education and Training management. The implementation evaluation of the programme was the responsibility of an evaluation group; a separate department within the education and training organisation. This reflects the widely held view in the literature with respect to integrity and independence of evaluators (i.e. Madaus et al, 1996; Guba and Lincoln, 1989). The programme development and delivery teams, who were also part of the education and training organisation, reported to the programme steering committee (Figure 4-VII).

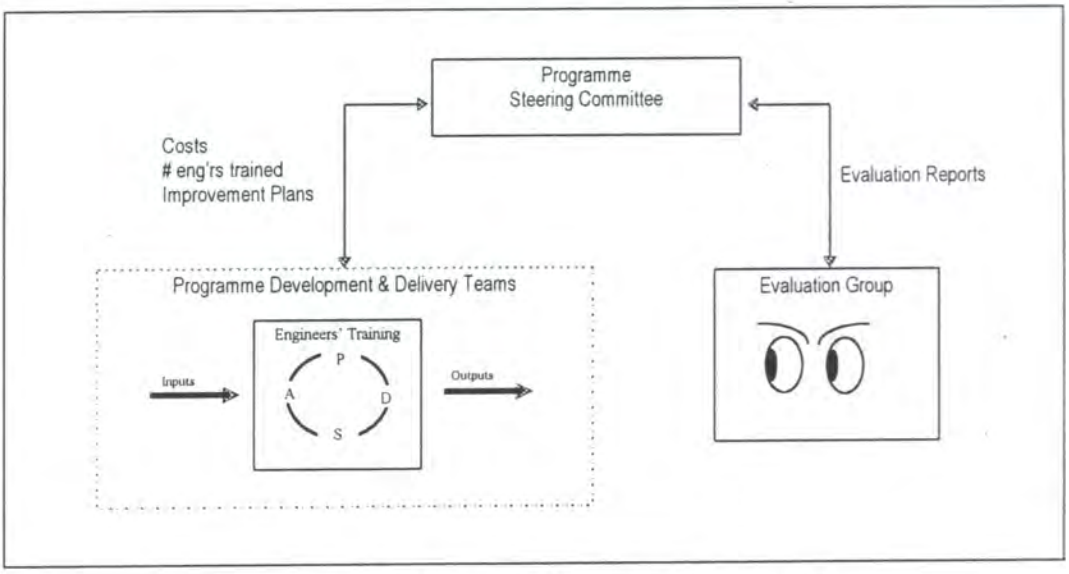

Figure 4-VII: Programme Management & Communication

Whilst these two groups contained some common members (e.g. myself and other members of the development and delivery teams), the two groups essentially worked independently from each other. Formal communication occurred primarily through the programme steering committee. The interests of these groups varied: The development and delivery teams, who had a close working relationship born out of the launch stages of the programme, as practitioners pursued continuous improvement of the programme in terms of its technical accuracy, feasibility of its content with respect to the engineering process, and the reactions of trainees. As such, they operated within a plan-do-study-act cycle of improvement. The evaluation group's interests lay in the development and application of reliable and valid methods of evaluation and providing periodic evaluation reports to the programme steering committee. The steering committee interfaced with the organisation at a senior level and as such were interested in efficient and cost effective management of the programme and demonstrating its contribution to the organisation.

4.6.1 Communication and Decision Making

Communication with the programme steering committee was facilitated by monthly meetings. The meeting were structured whereby key programme metrics were reported by the development and delivery teams. An established metric of training within the organisation was that of numbers of people trained with given resources and to this end the delivery teams were required to report monthly on the numbers of engineers trained per module and the utilisation of class places. Other metrics reported were incidences of non-attendance; development costs; and achievement of programme development timing objectives. Reports on the evaluation of the programme were made annually. The influence of evaluation information on programme decision making is considered in chapter 7.

The programme steering committee were responsible for strategic programme decision-making. The committee provided a forum for review of content and instructional design changes to the programme. Engineering managers influenced the programme's content through the committee.

4.7 Multiplicity of the Programme and Evaluation

From this comparatively brief overview of the evaluand content, training design and modular structure and implementation strategy, the reader will have gained an understanding of the diversity and complexity of the training programme. Whilst the programme was developed from a practical, as opposed theoretical foundation, many of its design concepts been related to the literature concerning learning and training.

With respect to evaluation and to this study, training is a multiplicity of variables which contribute to learning. Each activity in the conception, design, development and implementation of training is of interest to the training organisation. Many of the activities are of interest to other areas of an organisation, including trainees and line managers. Each activity requires managing and provides an opportunity for learning about training and improvement of training. Each activity therefore provides an evaluation opportunity.

Chapter 5

Programme Evaluation

5. Programme Evaluation

If you don't learn from your mistakes, there's no point in making them.

- Dr Peter Honey (1997)

Whilst evaluation of the programme had featured in the minds of the designers from a very early stage in its conception, no real thought was given to the subject until the programme was piloted in Sept 1990. A simple evaluation study was conducted with the pilot programme participants. The pilot evaluation was not systematically planned and consisted primarily of spontaneous data collection methods.

Following the pilot study, no further effort was put into evaluation until a few days before the Programme Foundation module was presented to the first of the two Technical Executive Groups as part of an approval process. For these courses, the development team produced a short pre / post training pencil and paper test as a means of demonstrating to the group how much their knowledge about quality methods had increased by participating in the programme. This simple test was to later form the basis of the knowledge assessment aspect of the implementation evaluation of the programme.

Following senior management approval to implement the programme across Europe, it was decided that, as the programme was the largest single training initiative to be undertaken in the company, an evaluation study of the programme's implementation would be conducted. This represented a major step into the unknown for many of the programme's stakeholders.

5.1 Evaluation Overview

The evaluation of the programme was conducted in two main phases; (i) pilot evaluation; and (ii) implementation evaluation (Figure 5-I). Each of the phases consisted of several components which are the subject of the remainder of this chapter.

Programme Development and Launch	Evaluation Phases
1990: Design and development of the programme	
1991: Pilot programme	Phase1: Pilot Evaluation
1992: Programme approval and launch preparation	
1993: Programme implementation	Phase 2: Implementation Evaluation

Figure 5-I: Phases of Programme Evaluation

As the pilot formed part of the development process, the development team were responsible for conducting the evaluation. Responsibility for the implementation evaluation was given to a separate department in the organisation, but necessarily involved members of the development and delivery teams (including trainers and administrators), trainees and their management, although the degree of involvement varied considerably - an issue which is considered in chapter 7.

5.2 Pilot Evaluation Phase

A pilot programme was delivered by members of the development team and its purpose was to evaluate the content, design, and materials of the programme. The pilot course consisted of seven modules, with four of the modules at two levels - see Table 5-I.

Module #	Module Name	Duration (days)
1	Programme Foundation	5
2	Process Management (level 1) Process Management (level 2)	3 4
3	Team Orientated Problem Solving	4
4	Failure Mode & Effects Analysis	2
5	Experimentation (level 1) Experimentation (level 2)	2 4
6	Quality Engineering (level 1) Quality Engineering (level 2)	2 3
7	Customer Focused Engineering (level 1) Customer Focused Engineering (level 2)	2 5

Table 5-I: Pilot Curriculum

A group of twenty-six engineers, which were representative of the target population for the programme, was identified and released by their management team. Whilst the make-up of the pilot sample was largely dependent on; a) the individuals volunteering; and b) their managers agreeing to release them, the final group assembled consisted of UK and Germany based engineers from Product Development, Manufacturing Staffs and Plants and the Quality Office. The group also reflected the age and experience range of the intended target population.

5.2.1 Purpose of the Evaluation

As part of the development of the material, the pilot presented an opportunity to test the relevance of the content of the programme to the engineering process and to observe the instructional processes employed for the training modules.

The purpose of the evaluation was to obtain reaction feedback from multi-national / cross functional operational engineers on; (i) the programme content; (ii) the relevance of the content to the engineering process; (iii) the integrity of the linkages between technical skills and with people skills; (iv) the programme's potential to improve quality; and (v) the instructional design of the modules and modular structure.

5.2.2 Pilot Study Evaluation Components

The pilot was evaluated using two formative methods; observation and participant (spoken) feedback at the end of each training day and two summative methods; focus group feedback session and questionnaire. The evaluation methods are described below. Formative and summative terms used here are in the context of the pilot and not of the overall programme.

5.2.2.1 Observation

The programme was delivered by the development team. For each module, the teaching was conducted by the specialist mini-team, with members of the core team observing and providing process help where required. The observation process was largely informal, but focused on the instructional design of the module (sequence, timing, presentations, exercises and materials) and the content validity, as received by the participants.

Observation notes were recorded onto copies of the training materials and in notepads. These were subsequently reviewed by the development team and used to make changes to the programme.

5.2.2.2 Participant (spoken) feedback

As part of the closing session (warm down) of each day of the programme, participants were invited for their comments about the day. This was normally undertaken using a "pass the pen" exercise; an idea based on an ancient North American Indian custom of using a peace-pipe to facilitate discussion. The exercise involved passing a pen around the participant group assembled in a circle. The rule of the exercise was that only the person with the pen could speak, saying something about the way the training day had gone and making a statement about their feelings.

The comments made by participants during this exercise were heard, and sometimes retrospectively noted down, by the members of the delivery / development team. This feedback was used to provide further insights into observations made during the training.

5.2.2.3 Focus Group Feedback Session

All pilot participants were invited to a half day feedback session conducted after the final module of the programme had been completed. A group session was selected as it was a convenient way of obtaining feedback from many individuals and it was likely to give rise to synergistic interactions between individuals (Brown et al, 1989). Of the original twenty-six participants, four had dropped out due to work commitments and sixteen attended the feedback session.

The purpose of the session was to identify aspects of the programme which worked well and aspects which needed to be improved. A two stage process for the session was employed comprising of a group brainstorm and discussion and completion of a questionnaire.

The group brainstorm process was facilitated using a process known colloquially as 'Metaplan'. The Metaplan process uses large portable pin-boards (measuring approx. 1.4 metres wide by 1.6 metres high) and cards of various shapes and colours. Information is written onto the cards and the cards are pinned onto the boards. Each card contains one idea and the cards can be moved around to indicate relationships to other cards. The cards can be written by the facilitator and/or the participants.

The Metaplan process has the advantage of collecting ideas very quickly in a manageable format for discussion and analysis. Ideas are generated by posing a question to which participants respond, either by calling out ideas (call-up question) which are written on the cards by the facilitator, or by writing their own cards (card question) which are collected and pinned onto the board. In each case, the group identify affinities between the cards which determine their relative position on the board.

For the feedback exercise, a combination of four card and call-up questions were used in the following sequence; (1) CARD: 'Identify 3 aspects of the programme which worked well'; (2) CALL-UP: 'Why?'; (3) CARD: 'Identify 3 aspects of the programme which could be improved'; (4) CALL-UP: 'How?'

The participants were randomly split into four groups. Each group was asked to respond to the four questions in turn. For each of the card questions, the ideas generated were discussed and the cards put into affinity groups. Cards duplicating the same idea were taken from the board. Then for each of the cluster groups the associated call-up question was asked, with participant responses written on cards and put on the appropriate position on the board.

The session was intended to generate qualitative data of the perceptions of the pilot participants. The structure of the questions and the nature of the metaplan process facilitated data analysis. The data was distinguished into positive and negative comments and through the process general themes emerged from correlated data items. These general themes identified the programme's areas of strength and weakness with individual data items identifying aspects for further consideration.

5.2.2.4 Feedback Questionnaire

A feedback questionnaire (Appendix D) was distributed to all respondents during the feedback session. The questionnaires were completed and returned by the respondents before they left the session.

The objectives of the questionnaire were to:

1. Assess the balance of the programme in terms of a) the content mix, b) the use of presentation (lectures) and syndicate work (tutorials), and c) the balance between conceptual knowledge and application training.

2. Assess the relative short and long term importance of various aspects of the programme.

3. Assess each module in terms of its content, relevance to the engineering process, course delivery (instruction), training materials, case study, and application to the workplace.

The questionnaire was designed to explore areas of the programme which were important to the development team in that they were unique to the programme or were considered to be important with respect to the programme's potential to bring about change in the engineering process. The questionnaire was administered anonymously to eighteen pilot participants as part of the focus group session.

5.3 Implementation Evaluation Phase

The main phase of the evaluation effort was concerned with the implementation of the programme and represents the bulk of effort and cost which was invested in evaluating the programme. This phase of the evaluation was initiated during the final stages of the development of the programme and in this sense was not fully integrated into the overall programme, but reflected much of the literature (e.g. Bramley, 1991) which conceptualise training evaluation as another step in the training process.

Following the pilot programme and the subsequent restructuring and changes to the programme instructional design, the programme was approved for full implementation by senior management. It was at this approval stage that the decision was taken to evaluate the implementation of the programme.

5.3.1 Purpose of the evaluation

The purpose of the implementation evaluation was to evaluate the effectiveness of the programme in terms of the programme's aims as stated in the Programme Foundation participant's manual - see Figure 5-II.

Programme Aims
The aims of the engineering quality improvement programme are to give engineers: • in-depth quality improvement skills appropriate to their function • an appreciation of what skills are appropriate to other engineers in different functions and some understanding of these skills • an understanding of their roles and responsibilities within the engineering process and how they can directly influence and improve the quality of the final product or service • skills which support all aspects of Total Quality Excellence • An appreciation of the linkages between the quality improvement skills

Figure 5-II: Programme Aims

Based on these aim statements, four fundamental evaluation questions were developed which, if answered, would indicate whether the programme had been successful. These four fundamental questions are given in Table 5-II.

Fundamental Questions	Outcome Characteristics
1. Has the programme changed the way in which Engineers think about quality	Engineers' knowledge and attitude changes
2. Have the changes in how engineers think about quality been maintained and transferred back to the work environment 3. Have the engineers behaviours actually changed	Engineers' behaviour changes
4. Has the programme been successful as an organisational change programme.	Organisational Changes

Table 5-II: Implementation Evaluation Fundamental Questions

The first question is concerned with two outcome characteristics of the programme: a) the level of knowledge engineers' would gain about the content of the programme, and b) the engineers' attitude towards quality as a result of attending the programme. Questions 2 and 3 are concerned with the retention and transfer of knowledge and attitudes gained and their application to the workplace in terms of changes in behaviour by engineers in approaching their work. The fourth question is concerned with changes to the organisation.

5.3.2 Evaluation Framework

The evaluation was modelled largely on Kirkpatrick's framework (1959a, 1959b, 1960a, 1960b). The framework was recognised amongst the training community within the company, although very little had been done beyond reaction level evaluation. As previously outlined, Kirkpatrick had featured prominently in training journals and other literature circulating the company.

Kirkpatrick's levels had been adopted within the company's training community to such an extent that it had become part of the language. When referring to evaluation, members of the training community would almost exclusively refer to level 1 evaluation; level 2 evaluation, as opposed to describing the evaluation of a programme in terms of the participants reaction at the end of a course or their performance in a knowledge assessment. The use of this language, particularly with respect to level 1: reactions type evaluation, added a virtual pseudo-scientific credibility to what was often otherwise a simple check of whether participants of a course had enjoyed themselves and felt they had made a good use of their time (Brittle, 1995).

At level 1 (Reactions), trainee reactions to the programme were obtained using a verbal feedback process which had been used as part of the pilot programme.

At level 2 (Learning), a knowledge assessment questionnaire methodology was developed and applied to assess changes in knowledge.

At level 3 (Behaviour), an attitudinal questionnaire methodology and behaviour observation methodologies were developed and applied.

At level 4 (Results), organisational analysis interview methodology was developed and applied.

Given the strands of some of the critical analysis of the literature in chapter 2 and the reservations contained there in, conclusions on the utility and success of using this framework are discussed in Chapter 7.

5.3.3 Application of the evaluation Components

The level 1 methodologies were integrated into the design of all modules and were therefore completed by every participant. The verbal feedback process was a feature of the original design as part of the people skills aspect to the training (see warm-down in Chapter 2). The questionnaire was included initially, but was subsequently dropped from the programme as described later.

The level 2 knowledge change assessment questionnaires were also integrated into the design of each module and were therefore completed by all participants.

As the above methodologies were integral part of the training course, they were administered by the programme trainer consultants.

The level 3 and level 4 methodologies were applied to a sample of the target population and stakeholder group. These were administered by external consultants to the programme.

5.3.4 Reactions to the Programme

In terms of Kirkpatrick, the level 1 type of evaluation; the participants reaction to the programme, this is probably the most easily observed. This was especially true for this training programme, where participants were encouraged in a structured way to voice their reactions to each of the modules. This provided direct feedback to the trainer at the end of each day, allowing the trainer to identify areas of concern amongst the participants. These concerns could range from the temperature in the room to instructional issues and dynamics of the group. The trainer could also further explore any comments made as part of the end of day / course review to gain a better insight into concerns raised.

Amongst the programme development team, a common view was shared about the value of collecting level 1 type of feedback in a documented way. Extensive efforts had been made, and were continuing to be made to improve all aspects of the instructional design. The modules were time constrained and the decision had previously been taken (at the approval stage) that a pre and post knowledge assessment would be included. Adding further questions to explore the participants reactions would only add time to the course agenda and duplicate much of the feedback which was received as part of the end of day spoken review. This evaluation was therefore not documented.

5.3.5 Measuring Changes in Knowledge

The knowledge assessment was concerned with the full range of the content of the programme modules, which included both the technical and behavioural skills. In this context, knowledge is defined as information (facts, principles, concepts) which can be recalled from memory when required.

The basic methodology used was a pre and post training knowledge level test administered at the start of each module and on completion of each module. As the participants completed and returned the questionnaire before leaving the course, the questionnaires were completed by all participants, with very few exceptions.

5.3.5.1 Questionnaire Design and Development

The original pre and post questionnaires were identical to each other for each module. They were developed by each module development mini-teams, which by this time had expanded to include the trainer-consultants on a voluntary basis.

The questionnaire design for each module was based on that for the Programme Foundation module, with the following basic design requirements; (i) they could be administered to participants in less than 30 minutes (at the start and end of the module); (ii) they could be marked and analysed in the classroom by the trainer consultant in less than 20 minutes; (iii) they would assess the range of technical and people skills content of the module.

The original Programme Foundation module pre and post questionnaire was developed by the core team. Given the time constraints, both in terms of time to prepare and to administer, 25 multiple choice questions were brainstormed. The question format consisted of a stem which took the form of a direct question or an incomplete statement and six possible responses. The possible responses for each question consisted of one 'Don't Know' response and five plausible responses, of which up to 4 could be correct, with the remainder being incorrect responses (or distracters).

The questionnaires were not piloted prior to launch and were not, therefore, subject to any validity or reliability assessment.

5.3.5.2 Application

The pre and post course questionnaires were administered as part of the course by the trainer-consultants. The pre-module questionnaire was administered to all participants shortly after arrival to the module.

Participants were given 30 minutes and asked to attempt the questionnaire using the multiple choice answer sheets in silence. They were instructed not to confer with colleagues and were supervised by at least one of the trainers present. On completion, participants retained a copy of their answers and returned the questionnaire question form and answer sheet by placing it in a box. Once all questionnaires and answer sheets had been returned, the answer and question sheets were separated and the answer sheets were marked by one of the trainers using a template which identified correct responses. For each group, their combined pre and post score data was presented to all participants using a histogram display.

5.3.5.3 Participant Identification

Due to the attitudes and established working practices in the company across Europe, participants would remain anonymous with regard to their performance in the tests. As the comparative analysis method selected required each individual's pre and post test scores to be compared against each other, a participant coding system was devised.

The coding system comprised of a number consisting of three components; Foundation module course number, participant number and year of course. E.g. F07-14-4 provides the following information; 'F07' refers to the foundation module number during that year; '14' is the unique participant number within the module; and '4' indicates the year of the module (1994).

Each participant was randomly issued with an identifier number, known only to the individual. Participants were asked to retain their number through subsequent modules. This allowed for knowledge gains of individuals through the modules of the programme to be studied and eleviated concerns among some participants that their training performance would be reported to management.

5.3.5.4 Mid-term Review

The initial pre and post knowledge questionnaires were identified to have two significant inherent design weaknesses.

1) Reliability: The first weakness identified concerned the reliability of the methodology to accurately indicate changes in actual knowledge levels. This was due to unquantifiable memory or practice effects on the recorded post scores and variations in chance probability of selecting the right response.

2) Validity: It was also observed that the original measure of knowledge did not fully cover all of the learning objectives of the programme and therefore did not measure was it was intended to measure.

5.3.5.5 Methodology Reliability Issues

With regard to memory or practice, as the same questionnaires were used for the pre and post module knowledge level assessment, any differences in the scores recorded for the post questionnaire, compared to the pre questionnaire, could be due to either changes in actual knowledge levels or as a result of memory and practice from completing the pre questionnaire.

Regarding chance probability of randomly selecting the right response, as there were multiple correct responses for each question and not a single correct answer, each question had a different probability of being correct. By answering every question by selecting all possible responses, a score of 100% would be gained using the scoring rules which had been established, as there was no negative marking scheme i.e., losing marks for getting a wrong answer.

A simple analysis of a sample of programme foundation modules was undertaken and revealed that although participants were scoring higher on the post tests, compared with the pre tests, they were also making more mistakes. The method and results of this analysis are described in the following section.

5.3.5.6 Comparative analysis of alternative marking strategies

The purpose of this study exercise was to analyse pre and post module questionnaire responses using alternative marking strategies and compare these interpretations of the questionnaire results with the results interpreted from the application of the current marking strategy.

A sample of 5 consecutive UK Programme Foundation modules; F07/94, F08/94, F09/94, F10/94 & F11/94 was taken. These modules had been delivered by a variety of (accredited) trainer-consultants. Alternative marking strategies were devised to provide the basis for comparison of the pre and post questionnaire responses. These were:

A. Incorrect responses only, disregarding correct responses. The number of incorrect responses are counted for each participant. Maximum number of possible incorrect responses = 73

B. Completely correct set of responses for each question. One mark is awarded for each question where all correct responses are given with no omission and no incorrect responses. Where an incorrect response is given or a correct response is missed, then no mark is awarded for the question. Maximum possible score = 25

C. Number of 'don't know' responses, regardless of correct or incorrect responses. Maximum number of 'don't know' responses = 25

For the selected sample, each participant's pre and post course questionnaire response sheets were scored using the alternative strategies. A summary of the scores, indicating the mean values and the range (standard deviation) of the pre and post tests for the samples using each of the alternative marking strategies are given in Figure 5-III, Figure 5-IV, and Figure 5-V.

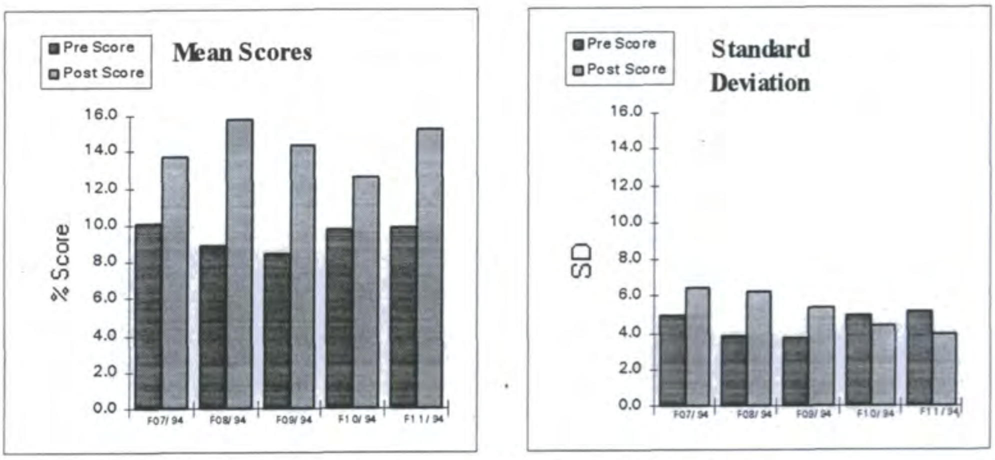

Figure 5-III: Alternative Marking Strategy A data

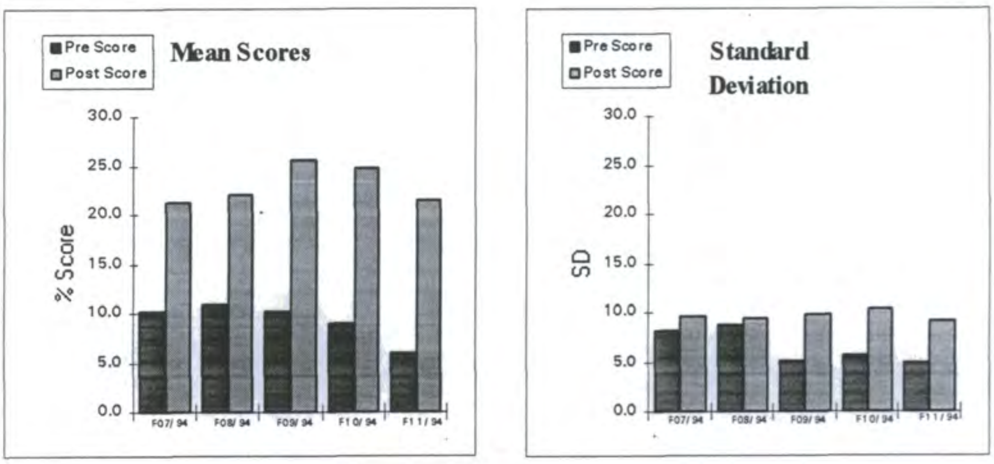

Figure 5-IV: Alternative Marking Strategy B data

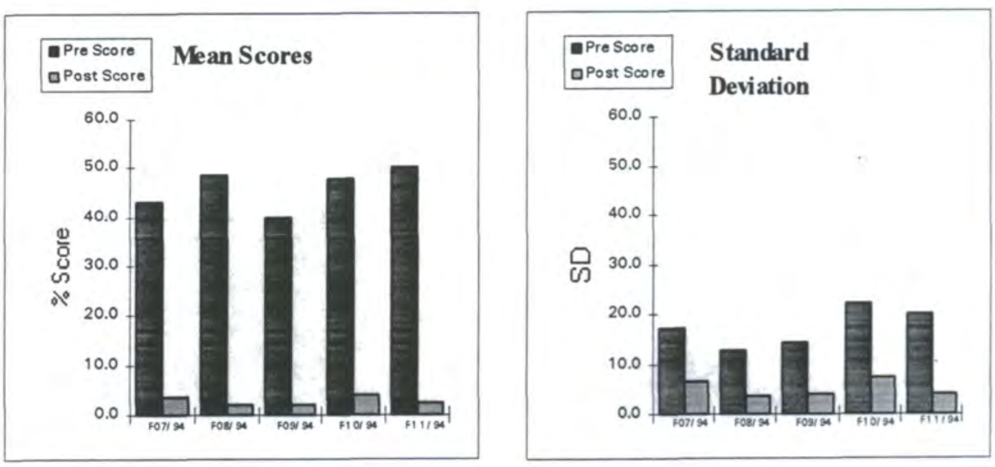

Figure 5-V: Alternative Marking Strategy C data

From the data it can be seen that for alternative marking strategy A (Figure 5-III), participants consistently score higher in the post training questionnaire. This indicates that participants answer more questions incorrectly after training than before. For alternative marking strategy B (Figure 5-IV), participants' scores increased for the number of completely correct responses to the questionnaire after training when compared with their pre-training score.

Finally, alternative marking strategy C (Figure 5-V) indicates a substantial reduction in 'don't know' responses by participants in the post training questionnaire when compared with the pre-training questionnaire 'don't know' responses. The reduction in 'don't know' responses and the correlating increase in incorrect responses suggest that trainees believe they know more than they actually do after training.

The increase in completely correct responses following training supports the conclusions from the original marking strategy although it is noted that post-training questionnaire means scores for both these strategies are not recorded above 50%. This suggests a high degree of difficulty of the questionnaire.

On the basis of this analysis, a review of the pre and post questionnaire design was undertaken. As part of the review parallel forms of the test were considered in order reduce the effect of participants attempting the same questions for pre and post knowledge assessment.

5.3.5.7 Methodology Validity Issues

Comparing the objectives of the modules against the pre / post questions which had been developed, it was observed (Brittle, 1994) that there was not a uniform correlation between the two. This had not been a consideration in the initial development of the questionnaires and so the development team had no way of knowing whether the knowledge being assessed were important facts, principles or concepts, or whether the feedback data was representative of the entire range of content of the modules.

The problem was compounded by the absence of well documented and up to date learning objectives for each of the modules. To examine the extent of the problem, a set of learning objectives were developed for the Foundation module using Bloom's taxonomy of cognitive behavioural objectives.

5.3.5.8 Programme Foundation: Key Learning Objectives

Learning objectives were developed from the instructional design (Table 5-III and Table 5-IV) and structured in terms of; programme philosophy; technical elements; people elements; content linkages and training (category); and topic, including training exercises. For each topic and exercise, the objectives were identified using Bloom's cognitive level hierarchy (indicated in brackets in the table).

The objectives statements were developed by a group of Programme Foundation module trainers with a member of the development team. For each item of the module agenda, the instructional materials were reviewed and the following question was addressed.

> *"What key piece of information should a participant gain, or how should behaviour change, as a result of participating in this session?"*

To distinguish levels of cognitive ability as a result of participating in the training, Bloom's levels of cognition were employed; level 1. Knowledge; level 2. Comprehension; level 3. Application; level 4. Analysis; level 5. Synthesis; and level 6. Evaluation. The level number is indicated in brackets.

The reader will note that only levels 1 and 2 were used for this analysis. This reflects the purpose of the Programme Foundation module which is to provide engineers with a conceptual understanding of the programme content. The only exceptions to this were for the Basic Quality tools and some of the people skills, where participants were expects to be able to apply the skills during and on returning to their workplace.

Category	Topic	Objectives and level ( )
Programme Philosophy:	The programme as a part of TQE	Describe negative and positive quality and how it has affected the company's quality strategy (1 Recall the Kano Model (1) Recall the six elements of Total Quality Excellence (1) Explain the customer supplier chain (2) Recognise the role of people skills in quality methods (2) Explain the quality lever (2) Explain how and why quality is defined by the customer (2)
Technical Elements:	TOPS	Describe the context for using TOPS (2) Describe the 8 disciplines of TOPS (2) Explain the strategic role of TOPS (2)
	Process Management (Exercise)	Explain the basic principles of the Shewhart control chart (2) Apply descriptive feedback for team process (3)
	Process Management	Describe the 4 steps of process management (2) Explain macro and micro process modelling (2) Explain and distinguish control and capability (2) Distinguish common and special cause variation (2)
	FMEA	Describe the 6 aspects of FMEA (2) Explain why FMEA addresses only negative quality (2)
	FMEA Exercise	Explain why failure modes relate to purpose or function (2)
	Experimentation	Explain that experimentation is a structured approach for gaining knowledge (2) State that experimentation can be applied in the planning stage of PDSA (1) Explain how experimentation allows for multi-factor changes (2) explain how experimentation identifies significant effect factors 2)
	Quality Engineering	Describe quality in terms of variability of function (as opposed to variability of parts) (1) Explain the concept of noise and robustness (2) Express systems in terms of energy and energy transfer (2) Describe Quality Engineering methodology as a design optimisation approach (1)
	Quality Function Deployment	Describe the 5 phases of QFD (2) Describe and distinguish QFD and traditional design process (2) Interpret 'what' to 'how' house of quality (2) Describe how the Voc is deployed to process controls thro' QFD (2) Describe Substitute Quality Characteristics and their significance to the design process (2)

Table 5-III: Programme Foundation Module Objectives (Sheet 1)

Category	Topic	Objectives and level ( )
Technical Elements (continued):	Quality Function Deployment Exercise	Apply affinity diagram technique to customer wants (3) Identify sales opportunities (2) Interpret customer competitive survey data and engineering competitive assessment data (3) Give descriptive feedback for team process (3) Establish and maintain team roles (3) Recognise disadvantages of large team meetings (2)
	Basic Quality tools	Apply the basic quality tools to appropriate situations (3)
People Elements:	Warming Up / Down	Explain the process of Warming up and the purpose of each stage (2) Explain the process of warming down and the purpose of each stage (2) Demonstrate how Warming up and down improves meetings (3)
	Listening Skills	Recognise own inhibitors to listening (2) Listen to and restate a message (3) Build on ideas by stating likes and wish (3)
	Team Meetings	Explain task, maintenance and process (2) Apply task, maintenance and process practices to team meeting situations (3)
	Action Planning	Apply the rules for brainstorming to generate possible solutions to a problem (3) Construct possible solution proposal which is feasible with a desireable outcome (3) Develop proposal through restatement modified by likes and wish (3) Identify action steps through brainstorming and like & wish selection (3) Formulate what, who, (where,) when action plan (3)
	Corporate Team Game	Recognise how lack of trust, communication and cooperation impaires company performance (2) Recognise the importance of goal setting and risk taking to affect change (2)
Content Linkages:	Linkages	Recall linkages within and between technical and people skills (1)
	Customer Focussed Engineering	Explain how 7 basic quality tools, TOPS, Process Management techniques, FMEA, Experimentation, Quality Engineering disciplines, Quality Function Deployment are interrelated in the context of a customer focussed engineering approach. (2) Explain how team building skills, communication skills, innovation skills and implementation skills are interrelated in the context of a customer focussed engineering approach. (2) Explain the conceptual relationships between technical and people skills in the context of a customer focussed engineering approach. (2)
Training:	Programme Experience	Recall the 7 modules of the programme (1) Explain the role of the Foundation module (2) Explain the role of the CFE module (2) Explain the ideal attendance sequence of the modules (2) Explain the consultancy service provided by the trainer-consultants (2)

Table 5-IV: Programme Foundation Module Objectives (Sheet 2)

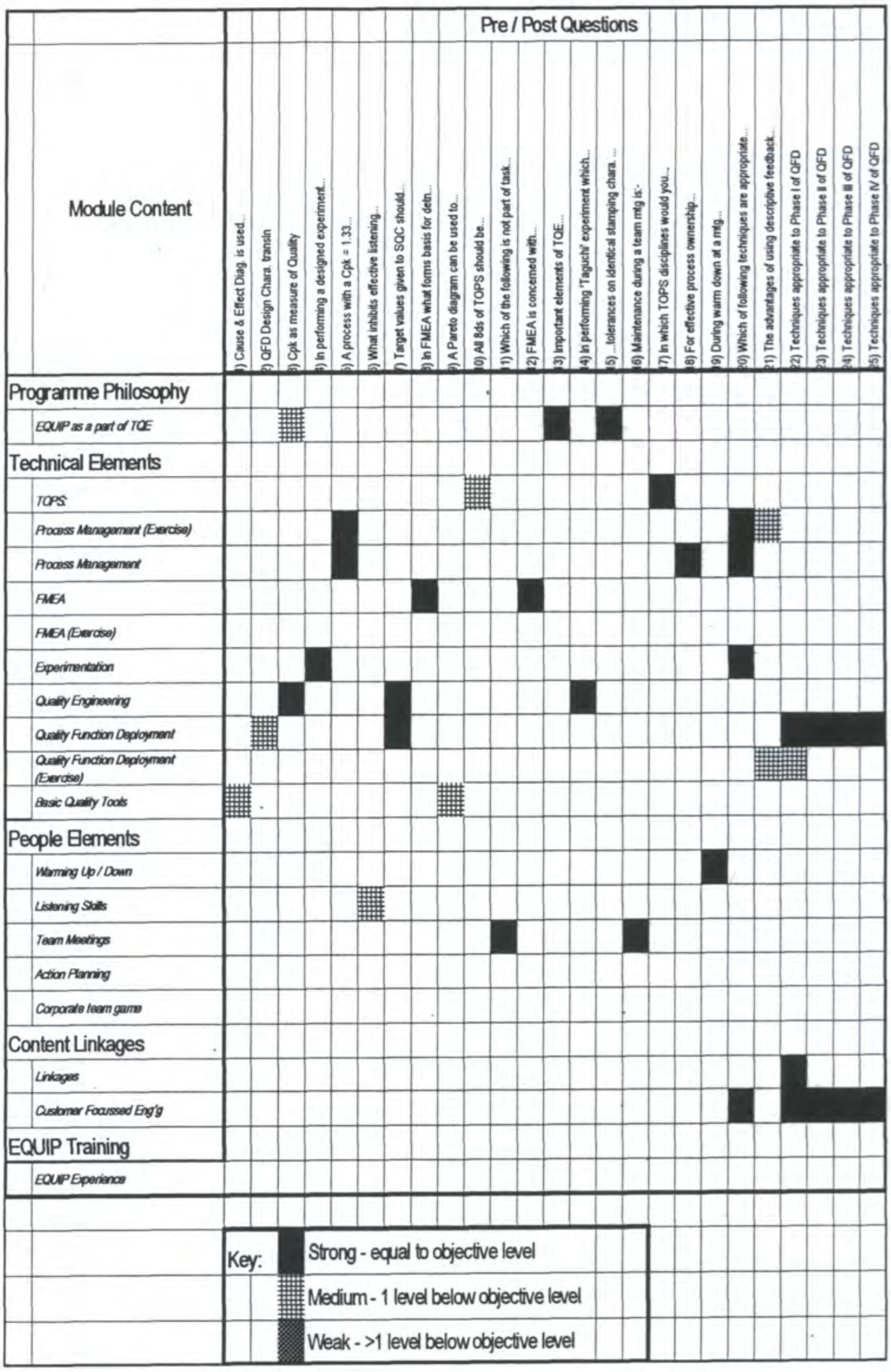

Figure 5-VI: Content/Question Comparison Matrix

Comparing the questions with the topic areas (Figure 5-VI) using a correlation matrix diagram, the correlation between the questions and the content of the module can be analysed. The strength of any correlation that exists is indicated using a 3 point scale based on Bloom's levels of cognition. Where the questions assess knowledge at the same level as stated in the objectives, a strong correlation exists, where the questions assess knowledge at one level below that stated in the objectives, a medium strength correlation exists and where the difference is greater than 1 level a weak correlation exists.

This scheme, whilst being highly subjective in its nature, particularly with regard to the strength of the correlation, provides an adequate basis for the assessment of the extent to which the pre and post knowledge questionnaires measure the range of content of the training. The questionnaire design was established with little prior experience of knowledge test development and was intended initially to serve the purpose of illustrating to senior management the knowledge adding potential of the training for engineers. It was quite literally developed by members of the core team in a few hours during the preparations for the senior management review.

Analysing the correlation matrix, blank rows indicate topic areas which are not assessed in the knowledge test. A blank column would indicate a redundant question, where the question cannot be tied to an item of the course.

Blank rows exists for the FMEA exercise, Action planning and the Corporate game and weak relationships exists for the QFD exercise, Basis quality tools, and Listening skills.

Defining validity as the extent to which a claim or conclusion is based on sound logic (Dane, 1990) or the degree to which an instrument measures what it is supposed or intended to measure (Oppenheim, 1992), the pre and post questionnaires were concluded to have low content validity.

5.3.5.9 Improved Knowledge Assessment

To address the reliability and validity issues outlined previously, the development team considered it necessary to develop an improved measure of content knowledge. The feasibility of applying the existing knowledge assessment was acceptable as overall format of a pre and post test administered during the module.

To improve the reliability of the knowledge assessment, a parallel form (Oppenheim, 1992) of assessment was adopted. A parallel form is where two measures are applied pre- and post-training. Both tap the same conceptual knowledge but use different questions, overcoming the issue of memory and practice effects. In addition only one answer which is correct for each multiple choice question is used. An essential element of the parallel forms is that although the questions are different, they are to the same degree of difficulty and address the same concept.

5.3.5.10 Development of the parallel knowledge measure

Development of the parallel forms of knowledge measurement was undertaken by a small team consisting of two trainers, and three members of the development team. The trainers acted as subject matter experts and the development team acted primarily as facilitators of the development process, although they also had a high level of knowledge about the content of the programme and its instructional design.

Twenty five multiple choice questions were developed for the Programme Foundation module knowledge assessment which required respondents to select 1 of the 5 possible multiple choice items as the correct answer. This gave a 20% probability of participants accurately guessing the correct response by chance.

With respect to validity, of the twenty five questions, nineteen questions were designed to assess the technical content of the module and six were designed to assess the people skills content of the module. The team agreed that the final set of questions addressed the important concepts across the range of topics in the module and the knowledge measure were the optimum that could be achieved given the application constraints which had been set.

The final questionnaires for the Programme Foundation module are given in Appendix E.

5.3.5.11 Pilot test

The parallel forms questionnaire was pilot tested and the data collected was analysed to establish the reliability of the measure. To confirm that the questionnaires are parallel, two properties were analysed; (i) that equivalent content knowledge (principles, concepts or information) is assessed in the pre and post test; and (ii) that the level of difficulty of the pre and post tests are equivalent.

The pilot was conducted with a sample of twenty-three engineers, who completed the pre and post tests without undergoing the Programme Foundation module. The questionnaires were scored using the new scoring procedure and the results were analysed in terms of the two properties being assessed.

5.3.5.12 Equivalent content knowledge

To establish whether the equivalent content knowledge was actually being assessed by the pre and post measures, the degree of the relationship between the two sets of data was analysed using Pearson's correlation coefficient (r) (Pearson and Hartley, 1954) in terms of: i) all questions, ii) technical questions, and iii) the behavioural skills questions. Coefficients of correlation provide estimates of the relation between the pre and post questionnaire items. Kerlinger (1986) describes this approach as the most commonly used method for calculating indices of relation in behavioural research. Oppenheim (1992), referring to self consistency of a measurement instrument, notes that in social and behavioural sciences it is rare to find consistency of greater than 81% ($r > 0.81$) of common variance, or co-varaince.

The resultant coefficients and their statistical significance (p) are given in Table 5-V.

	All Questions	Tech. Questions	Behav. Questions
Pre-test Scores	r = 0.5	r = 0.68	r = 0.36
(Significance)	(p = < 0.01)	(p = < 0.0001)	(p = < 0.05)
	Post-test Scores		

Table 5-V: Equivalent content knowledge test correlation

The correlation coefficient (r) for technical questions was found to be 0.68, for behavioural skills questions was found to be 0.36, and for the overall content (all questions) assessed it was found to be 0.5. The significance of these are <0.01, <0.0001 and <0.05 respectively. It was, therefore, concluded that the pre and post module knowledge test did assess equivalent content knowledge.

To analyse the relative levels of difficulty of the pre and post module knowledge questionnaires, the scores data collected for the two tests was compared in terms of all questions, analysed in terms of technical difficulty and behavioural skills difficulty (Table 5-VI).

	Pre Scores		Post Scores	
	Mean %	% SD	Mean	% SD
All Questions	21	10.106	24	11.6307
Technical Questions	18	13.522	20	13.367
Behavioural Questions	31	11.319	36	20.073

Table 5-VI: Level of difficulty comparison

The means scores (and standard deviations) for the pre and post measures do not significantly differ. In the absence of training, the scores for the pre and post tests are expected to be the same, if the level of difficulty of the questions is the same. Comparing all pre and all post questions, the combined scores (pre test mean of 21% and post-test mean of 24 %, and pre-test scores range (SD) of 10.106 and post-test scores range (SD) of 11.6307) it was concluded there was no significant difference in either the mean scores or the range. Further, the analysis indicated that participants were scoring at a chance level of 21% and 24 % for the tests, which given the design of the questionnaire of one correct response out of five options (20%) chance, is not significantly different.

Comparing combined scores for the technical questions, there was no significant difference in mean scores or range and participants were scoring at chance level (20%+). For the behavioural skills questions, there was no significant difference in mean scores or range, however, participants were scoring at above chance level (31% and 35% respectively). It was decided that this did not warrant further development of the questionnaire in view of the time and effort required to redesign and revalidate.

5.3.6 Measuring Changes in Attitudes

Attitudes play a central role in training and change. They are the favourable and unfavourable reactions to objects, people, situations, or any other aspects of the world, including abstract ideas and social policies (Atkinson et al, 1993). Attitudes are linked to thinking and behaviour and therefore attitudinal aspects of training are extremely important as they predispose learners to action (Reid and Barrington, 1994).

A stated aim of the programme was to 'change the way engineers think'. This statement was made in the context of quality improvement and so as part of the evaluation of the implementation of the programme, changes in attitudes of participants were assessed.

The attitudes assessment focused on two aspects of engineer's thinking: i) how they perceive their organisation, and ii) how engineers think about the concept of quality. The perception of the organisation is significant in two respects; firstly, the importance of the role of the organisational climate in the improvement of quality is well documented in the literature (i.e. Dale, 1990; Juran, 1989; Oakland, 1993); and secondly, Broad and Newstrom (1992) observe the significance of a supportive environment as perceived by the learner on the transfer of training. Reid and Barrington (1994) similarly observe that the climate of the organisation as a powerful influence in determining whether training is likely to be transferred to the working situation.

5.3.6.1 Measurement Development

Measuring changes in attitudes is difficult; attitudes are constructs in that they are abstract concepts which cannot be directly observed. Changes in attitude can only be inferred by a person's words and actions. Henerson et al (1987) identify four precautions for measuring attitudes; a) measuring attitudes relies on inference, since it is impossible to measure attitudes directly, b) behaviours, beliefs, and feelings will not always match, so focusing on one manifestation of an attitude may tend to distort and mislead, c) there is no guarantee that an attitude will not be volatile or fluctuate for a one time measurement to be reliable, and d) there may not be a universal agreement to the nature of an attitude which is the subject of measurement.

For the purposes of the evaluation, participants' attitudes were inferred using questionnaires and attitude rating scales. Two types of rating scale were used; i) normative (Likert) scale; and ii) Ipsative scale.

The distinction between normative and ipsative scales of measurement is generally misunderstood in research and measurement (Kerlinger, 1986). Normative measures are measures which vary independently and as such, they are relatively unaffected by other measures. They are used for interpretation to the mean of the measures of a group (individuals' sets of measures having different means and standard deviations). E.g: if a scale of 1 to 5 (where 1 is poor and 5 is excellent) is available to rate the quality of 5 vehicles, the assessor will assess each vehicle and assign a rating. If all vehicles are of excellent quality, then the assessor will rate them all 5. This is an example of a normative measure of quality.

Ipsative measures are systematically affected by other measures and are used for interpretation to the same mean (each individual's set of measures having the same mean and standard deviation). E.g. if a scale of 1 to 5 (where 1 is the highest quality and 5 the lowest quality) is available to an assessor to rank 5 vehicles in terms of their respective levels of quality, then the assessor must decide which vehicle has the highest quality level of all the vehicles and assign the ranking (1), then the next highest quality level vehicle and assign the ranking (2) and so on through to the last vehicle which is of the lowest quality level of the five and assign the ranking of 5.

To develop the measures, a team of six programme and other trainers with considerable experience were assembled to identify questionnaire items which would indicate how participant' perceptions of their organisation and of quality. The team met over a period of several months to identify and refine the questionnaire items into a measurement system. The process was facilitated by members of the evaluation team.

For participants' attitudes with respect to the organisation, a normative measure of their perceptions was developed using graphic format scales (Dane 1990) and applied. For participants' attitudes with respect to quality, normative and ipsative measures using graphic format and forced choice scales (Dane 1990) were developed and applied.

5.3.6.2 Development of 'perceptions of organisation' Measure

To develop the normative measure of participants' perceptions of the organisation, the team employed a brainstorming process to identify exemplars of a quality improvement supportive organisation which were then developed into statements to be used in conjunction with a nine point graphic format scale.

The exemplar statements were grouped in terms of their affinity to one another to form supportive organisation dimensions of; 'principled management'; 'security in help seeking'; 'management efficiency'; and 'trust in Colleagues' (Table 5-VII). The statements and rating scales were produced as a questionnaire which is given in appendix F.

Dimension	Questionnaire Item
Principled Management	Management is sincere in its attempts to meet the employees point of view
	There is a strong sense of community, a feeling of shared interest and purpose among the managers of the company.
	Our management would be quite prepared to gain advantage by deceiving the employees.
	I feel quite confident that the company will always treat me fairly
	There is little conflict between managers
	Meaningful co-operation and innovations in the company are stifled because of too many vested interests
	Managers are more interested in achieving the organisational goals than in personal advancement
	The company has a poor future unless it can attract better managers
Security in Help Seeking	Sometimes I worry that asking for help at work might look like I can't do my job.
	Asking for help from my colleagues can sometimes be humiliating
	I am wary of asking for help publicly at work
	It takes courage to ask for help in this organisation
	I have to be careful when I talk to colleagues about work difficulties
	Asking someone for help at work is as easy as asking a favour from a friend
	I am happy to admit it when I need help to do my work
	It is expected that one asks colleagues for help at work
Management Efficiency	Management at work seems to do an efficient job
	Management can be trusted to make sensible decisions for the future of the company
	There is considerably more competition than co-operation among the managers in the company.
Trust in Colleagues	I have full confidence in the skills of my colleagues
	Most of my colleagues can be relied upon to do as they say they will do
	I can rely on my colleagues not to make my job more difficult by careless work

Table 5-VII: Perception of Organisation Dimensions

The rating scales were organised so that high scores on the Principled Management dimension would indicate trust in management and that the participant views them as being co-operative. High scores on the Security in Help Seeking dimension would indicate that the work environment is very supporting to seeking help. High scores on the Management Efficiency dimension would indicate that the participant has trust in management and considers them to work efficiently. Finally, high scores on Trust in Colleagues would indicate that the participant places trust in the skills and reliability of colleagues.

5.3.6.3 Development of 'perceptions of quality' Measure

Perceptions of quality is a very general notion incorporating a diverse range of concepts. To define this overall measure, the team sub-divided perceptions of quality in terms of; innovation and excitement; customer empathy; creative engineering; product innovation; and open to new approaches.

Innovation and excitement reflects the shift in emphasis from quality being concerned with simply meeting customer needs to exceeding their expectations of the product or product feature. This is conceived in the programme as the notion of positive quality, where quality is defined in terms of excitement features and as such is largely dependent on engineering innovation (chapter 4). As the measure is concerned with a traditional and new notion of quality, a forced choice item scale was selected.

The team identified a range of concepts which were associated with innovation and excitement. These were refined into seven questionnaire items by identifying discrete and mutually exclusive options for each. E.g. for the concept of positive and negative quality, the options developed are; (A) To improve product quality, more effort should be directed towards preventing problems or (B) To improve product quality, more effort should be directed towards creating innovative products

Option (A) reflects an approach to improving negative quality, and (B) reflects an approach to improving positive quality. Both are related to quality and are important, but option B exemplifies the shift in attitudes towards positive quality intended by the programme.

The innovation and excitement items are given in Table 5-VIII which were produced as a questionnaire which is given in appendix G.

Questionnaire Item
A. To improve product quality, more effort should be directed towards preventing problems B. To improve product quality, more effort should be directed towards creating innovative products
A. To improve product quality more effort should be directed towards solving problems B. To improve product quality more effort should be directed towards creating innovative products
A. Quality is about exciting the customer B. Quality is about preventing mistakes
A. Problem solving should be recognised and rewarded in the same way as improvements B. Innovations should be recognised and rewarded in the same way as problem solving
A. Quality products must have new and exciting features B. Customers prefer new products to be reassuringly familiar
A. The quality of a product or service depends entirely on our technical ability B. To make significant improvements in quality you need to be creative in your solutions
A. An effective measure of quality is a customer satisfaction index B. An effective measure of quality is the number of customer complaints

Table 5-VIII: Innovation and Excitement

High scores on this scale would indicate that the participant's perception of quality is in terms of excitement features to customers and therefore has recognised the significance of the notion of positive quality. Low scores would indicate that the participant continues to view quality in traditional terms.

Customer empathy, creative engineering, product innovation, and open to new approaches reflect the wider range of concepts within the programme. To develop these measures, the team brainstormed associated statements which were developed into 15 items (Table 5-IX) using a nine point disagree/agree graphic format scale. E.g. as an indicator of open to new approaches, the statement 'once you have a system, it is best to stick to it' can be rated as '1' (strongly disagree) or '9' (strongly agree) or somewhere in between. 'Strongly agree' reflects an attitude for trying new ways of engineering, whereas 'strongly disagree' indicates an attitude of resistance.

These items which were produced as a questionnaire which is given in appendix H.

Dimension	Item
Customer Empathy	Customers don't know what they want until experts show them
	Sometimes we should give people what we think they need, and not what they say they want
	The customer is very often wrong
	When evaluating quality it is always better to consider the opinions of internal specialists rather than external clients
	My contribution to the overall product is negligible
	Most new features on our products are merely gimmicks
Creative Engineering	It is important to take time out to develop new and exciting concepts
	To make significant improvements in quality you need to be creative in your solutions
	I ought to have a sound understanding of the skills in other engineering areas
Product Innovation	A quality product must have new and exciting features
	Customers prefer new products to be reassuringly familiar
	A good way to redesign a new product is to adapt the old design
Open to New Approaches	Once you have a system, it is best to stick to it
	Customer satisfaction is all that matters
	When solving problems, I prefer to start with traditional approaches

Table 5-IX: Perception of Quality

High scores on Customer empathy factor would indicate that respondents value the input from customers in terms of the engineering process. High scores on creative engineering factor would indicate a desire to strive for creative solutions whilst acknowledging the need to understand as much as possible about all areas of engineering. For the product innovation factor, high scores would indicate that participants perceive the development of new and exciting ideas as a contributing to high quality products. Finally, high scores on the open to new approaches factor would indicate that participants regard quality as being developed by adopting new approaches, as opposed to using traditional techniques.

5.3.6.4 Reliability and Validity for the Attitude Scales

The reliability of the measures was determined from the final survey data in terms of each item's correlation to its dimension. Pearson's product-moment correlation of coefficient was used and items with values greater than 0.7 were accepted to be reliable. Items which were found to be unreliable have been excluded from the measures.

The measures were assumed to have content validity in that they were developed by subject matter experts. To assess the concurrent validity of the measures (how well the tests correlate to each other), the correlations between the dimensions were examined and it was concluded that internal validity existed between these measures. High scores on the total (ipsative) creativity and innovation scale, correlate to high scores on total perceptions of the organisation scale and on the total normative perceptions quality scale. This indicates that participants perceive quality as involving creativity and innovation, and an organisation that is supportive and co-operative. It also suggests that quality involves innovative problem solving.

5.3.6.5 Application of the Measures

Although the questionnaires were administered to a total of 983 engineers (Table 5-X), this was done at 3 stages. Although a high number of engineers completed the questionnaire at stage 1, for stages 2 and 3, a smaller number of engineers completed the questionnaire. This was due largely to changes in administrative organisation and the level of buy-in, or commitment, to the evaluation by the wider training implementation team. This issue is considered in chapter 7 as part of the meta-evaluation of the study.

Stage in Programme	Number of Respondents
Stage 1: Prior to the programme, where no training had been received.	945
Stage 2: After completing 3 training modules	117
Stage 3: After completing all level I modules	120

Table 5-X: Administration of Attitude Questionnaires

5.3.7 Measuring Changes in Behaviour

As part of the evaluation of the implementation of the programme, measures were developed to assess how successfully the engineers' applied the skills acquired through participating in the training modules of the programme on their return to the work place.

Initially, two approaches were identified with one intended to measure the technical skills and the other to measure the people skills.

5.3.7.1 Application Checklists.

Intended originally as a training aid, application checklists were developed for all modules to serve two main purposes; (i) to provide participants with a checklist of stages in the application of the technical skills to the engineering process; and (ii) to provide engineering managers with an understanding of what was involved in applying the methods, in terms of engineering time on the part of the participants, so that they could actively support the applications.

In their initial form, the application check sheets, provided the basis for development of the measure. Due, in the main, to time constraints, however these were not developed. Although the check sheets were applied by trainer-consults and, in all probability, by some participants, records of applications were kept. No evidence could therefore be given as to the application of the technical skills. A major deficiency of the evaluation which is discussed in chapter 7.

5.3.7.2 Behaviour Rating Scales

To measure the application of the people skills of the programme, behavioural rating scales were developed. The scales were developed for multiple purposes; i) to evaluate the programme, ii) as an instructional aid for use by trainers to measure performance of participants during syndicate exercises, iii) as a diagnostic tool for use by the trainer-consultants to assess the performance of work-based teams, and iv) by members of work-based teams to measure their own team performance.

5.3.7.3 Development Process

To develop the scales, a team consisting of programme trainer-consultants and a process facilitator participated in a series of workshops designed to elicit behavioural descriptions of a range of successful, average and unsuccessful applications of the people skills. The team selected nine team-building and communication skills to be included in the measure; warm-up; team roles; task and process review; warm-down; listening; questioning, facilitation, effective speaking; and feedback.

The team employed critical incident technique to identify behavioural descriptions reflecting the range of applications from good to bad. The data generated during the sessions was then analysed and collated to develop categories of incidents. A process of review and revision by the team refined the descriptions, before agreeing performance dimensions.

The behavioural descriptions were then given to the whole group of trainer-consultants who were instructed to retranslate and verify the behavioural descriptions as being exemplars of each behavioural description. This allowed for the content validity of the scales to be checked as the trainer-consultant group had considerable knowledge of training and applying the people skills.

In addition to retranslating the descriptions, this group of participants were also asked to rate the behaviour described as to how effectively/ ineffectively it represented the performance on the appropriate dimension. These ratings were based on a seven point scale (1 = poor performance to 7 = excellent performance). The mean, mode, median ratings and the standard error of measurement (SEM) for each description was calculated from the participant responses to ascertain their degree of agreement for each. The resultant scales are given in Table 5-XI and Table 5-XII.

Warm Up	Mode	Std. Error
1. group start task immediately	1.0	0.20
2. purpose unclear to group members	2.0	0.34
3. purpose & task made clear but no attention to maintenance	3.0	0.55
4. purpose & task made clear & members are introduced to each other	4.0	0.20
5. purpose & task made clear & members	5.0	0.34
6. expectations were declared with a common understanding of task roles & task/time planning	6.0	0.26
7. attention paid to all 6 stages of warm-up	7.0	0.34
Team Roles		
1. no roles assigned	1.0	0.14
2. team members do not recognise roles assigned	2.0	0.34
3. use of focused questions & not carrying out assigned role	3.0	0.36
4. members assigned roles but don't stick to own roles/take on others'	4.0	0.20
5. roles assigned but evidence of misuse. E.g. scribe filtering, time checks as opposed to time management	5.0	0.28
6. roles clearly displayed during the session, leader, facilitator, scribe, time manager, team members	6.0	0.18
7. roles clearly displayed during the session & evidence of changing roles to suit changing circumstances within the team	7.0	0.14
Task		
1. no methodology identified to deal with the task	1.0	0.24
2. no structure, inconsistent in approach, no prioritisation	2.0	0.34
3. structure and agenda agreed but no time plan	3.0	0.28
4. time plan not carried out	4.0	0.31
5. agreed process, fully structured & consistent approach but task incomplete	5.0	0.18
6. agreed process, fully structured & consistent approach but extra time needed to complete task	6.0	0.34
7. an agreed process, fully structured & consistent approach but extra time needed to complete task, fully prioritised process & completed within the time plan	7.0	0.29
Process Review		
1. no review of meeting carried out	1.0	0.76
2. do review but use evaluative comments	2.0	0.51
3. review undocumented	3.0	0.45
4. review carried out taking into account observations from team members but not feelings	4.0	0.14
5. review carried out taking into account observations from team members & feelings	5.0	0.14
6. review carried out with clear sequence, & notice roles, feelings & changes	6.0	0.20
7. full review of meeting carried out & factors for changing clearly identified & fully documented	7.0	0.29
Maintenance		
1. straight into task & without checking whether people understand the methodology	1.0	0.18
2. straight into task and no account of people's feelings at the beginning	2.0	0.34
3. not checking out how individuals are feeling during meeting	3.0	0.48
4. recognising that someone hasn't contributed and not doing anything about it	4.0	0.20
5. evidence of checking out how individuals felt during meeting	5.0	0.48
6. ensuring that all team members are given an opportunity to express their feelings on specified issues	6.0	0.26
7. frequent maintenance checks to ensure that people fulfil the task	7.0	0.28
Warm Down		
1 participants leaving the meeting before the end without being acknowledged	1.0	0.00
2 meeting finishes when task ends, no review of task	2.0	0.26
3 no review of purpose	3.0	0.42
4 no opportunity to make closing statements	4.0	0.79
5 opportunity to make closing statements which include personal feelings (no "I feel that....")	5.0	0.14
6 group acknowledged each other's contributions	6.0	0.20
7 attention paid to all stages of warm-down	7.0	0.18

Table 5-XI: Team-building Scales

Listening Skills	Mode	Std. Error
1. other person making a statement disconnected to previous statement	1.0	0.20
2. use of negative responses e.g. "I hear what you say but...."	2.0	0.30
3. interrupting others	3.0	0.26
4. evidence of listening but no restatement	4.0	0.55
5. nodding, evidence of listening, use of "Yes, and...."	5.0	0.20
6. use of 'likes' and 'wishes'	6.0	0.20
7. restatement of original proposal to include 'wishes'	7.0	0.20
Questioning		
1. use of leading questions	1.0	0.28
2. extensive use of closed questions	2.0	0.36
3. narrow questions without gaining more information	3.0	0.29
4. questions used for clarification	4.0	0.47
5. varied intonation in delivery of question	5.0	0.00
6. use of relevant & focused questions	6.0	0.20
7. appropriate use of different types of questions in order to elicit relevant information	7.0	0.56
Facilitation		
1. recognise that someone is being ignored without doing anything about it	1.0	0.14
2. discussing items disassociated with the agenda	2.0	0.24
3. unrealistic vision of success	3.0	0.18
4. stated objectives which are difficult to measure	4.0	0.34
5. recognising there is a problem but fail to deal with it immediately	5.0	0.28
6. deliberate break to check everyone is happy with progress	6.0	0.54
7. evidence of facilitation (e.g. 'I notice you haven't contributed for some time, is there anything you would like to add on this matter?)	7.0	0.18
Framing Information		
1. lack of structure, no links, flow, disjointed, separate entities, no account taken of audience level	1.0	0.37
2. no headline, no summary, verbose statements, no distinction between subject areas	2.0	0.30
3. distinction between subject areas but not made explicit & full substance of text not emphasised	3.0	0.26
4. presentation gives an outline at the outset but some key points missed	4.0	0.47
5. uses links to emphasise relationships between different points	5.0	0.20
6. key points used to emphasise important information and pitched at audience level	6.0	0.34
7. structured presentation with evidence of headlines, signposts, key points & boundaries, links clearly targeted at the optimum audience level	7.0	0.26
Feedback		
1. no feedback given to the group	1.0	0.42
2. evaluative feedback using negative or emotive words	2.0	0.78
3. comments made without reference to the evidence	3.0	0.29
4. descriptive feedback given but group members were not allowed to respond	4.0	0.18
5. group members given the opportunity to express what they notice & how they feel	5.0	0.28
6. descriptive feedback used extensively with group given the opportunity as to what they would change for their next meeting	6.0	0.42
7. a positive change in group behaviour as a direct result	7.0	0.26

Table 5-XII: Communication Scales

This analysis indicates that for each of the People skills seven identifiable behaviours. These represent the degrees of good or bad behaviour for each of the people skills selected for the evaluation.

5.3.7.4 Reliability & Validity of the Scales

To assess the reliability of the scales, two observers used the scales to observe syndicate exercises in the FMEA and TOPS programme modules. The observation ratings of each were compared to establish inter-rater reliability, which was found to be high. The scales were assumed to be valid as they had been developed and checked by experienced trainer-consultants.

5.3.7.5 Application of the Scales

Observations of two teams applying the FMEA method to the engineering process was conducted using the developed scales. One team had undergone the training and the other had not. The observations were made by a member of the evaluation team who rated team performance for each of the selected people skills. The observer attended both team's FMEA meetings and their role was declared to teams. The observer sat at the back of the room and did not directly influence the process.

In addition, each member of both teams was asked to assess themselves against the scales at the end of the meeting. This provided a self-assessment of their performance. The resultant data is given in chapter 6.

5.3.8 Measuring Programme Effectiveness

The evaluation strategy employed stakeholder analysis to investigate the perceived success of the programme. Stakeholders are those people who have a stake in the outcome of an evaluation or programme (Patton, 1982), or are potential users of evaluation results and others who may be affected by them (Cummings, 1998). In this context, the role of the stakeholder is to provide feedback on the effectiveness of the programme itself, and not to provide input to how the programme should be evaluated as in the concept of a stakeholder approach to evaluation developed as part of the 'qualitative -naturalistic - descriptive methodology' of the responsive movement in the 1970s (House, 1986). The consequences of this on the utility of the evaluation are discussed in chapter 7.

Stakeholder analysis (Guba and Lincoln, 1989; Burgoyne, 1992; 1994) provides a methodology where the perceptions of the progress and effectiveness programme can be elicited and used to draw conclusions about the programme.

5.3.8.1 Stakeholder Identification

To develop the methodology, the stakeholder groups had to be defined and identified. Stakeholders were identified by members of the development and delivery team through a brainstorming process and narrowed down to those stakeholder groups described in Table 5-XIII.

As the major part of the programme's implementation was in the UK and Germany, and as the company's engineering centres were located in these two countries, individuals were identified for the stakeholder groups from these two countries.

	Number of Interviewees		
Stakeholder Groups	UK	Germany	Total
Senior Managers	3	0	3
Course Designers (Core Team)	3	0	3
Trainers/ Internal Consultants	8	8	16
Trainees	8	8	16
Managers of Trainees	8	8	16
Course Administrators/ Co-ordinators	1	0	1
Total	31	24	55

Table 5-XIII: Stakeholder Groups (1993)

For each stakeholder group, individuals were identified and data was collected from each individual using semi-structured interviews. A total of 55 stakeholder interviews were conducted (31 in the UK and 24 in Germany). Each interview lasted approximately one hour and was recorded using audio tapes. All interviewees were assured that the information collected would be treated in the strictest confidence.

5.3.8.2 Interview design and development

A semi-structured exploratory interview method was selected to ensure that the same question areas were covered consistently across each of the stakeholders (Oppenheim, 1992). The interviewee responses were then probed for further information. The interview themes were established by the team based on the stated aims of the programme.

To identify the interview questions areas, six trial interviews were conducted with members of the development and delivery team using two external interviewers. Given the organisational level of some of the stakeholders identified and the sensitive nature of the interview themes, it was decided that by using external interviewers, the data collected was less likely to be prone to interviewer / interviewee interaction errors (Mosser and Kalton, 1971).

The interviews were taped and transcribed. The transcriptions were content analysed (Krippendorff, 1981) by each interviewer to refine the themes into categories. The code-book was revised several times by both interviewers through a process of re-examining the transcripts. The resultant code-book is given in Appendix I. The pilot study enabled the assessment of the suitability of the question areas and the development of the coding system for the qualitative data analysis.

The question themes included in the final interviews were as follows:

1. **Involvement**: Interviewees were asked to describe their extent of involvement in the programme, what the success/ failure of the programme meant to them and how much control they perceived they had over the success of the programme.

2. **Objectives and Achievement**: Interviewees were asked what they thought the programme would accomplish and whether they thought that the company needed the programme. They were also asked to what extent they thought that the programme was relevant to an engineers daily work.

3. **Content**: Interviewees were asked how they thought the programme could be improved and how the programme related to the company's philosophy of Total Quality Excellence. Interviewees were also asked whether they thought the programme would change the way that engineers perceive quality.

4. **Impact on Customers/ Suppliers**: Interviewees were asked about the ways in which the programme would impact on other functions within the Company and how the programme might affect external suppliers.

5. **Barriers**: Interviewees were asked what they thought the (potential) barriers to the success of the programme would be.

6. **Organisational Climate:** Interviewees were asked to describe how they perceived the current organisational climate and whether they thought the programme would have any impact on it.

7. **Perception of Training Department**: Interviewees were asked to describe their perception of the training function within the company and whether the programme had altered their perception in any way.

8. **Future Vision**: Interviewees were asked what they envisaged was the future of the programme and what it will mean to the future of the company. Interviewees were also asked to comment on the success of the programme as a cross-cultural programme.

5.3.8.3 Reliability and Validity

The reliability for qualitative inquiry refers to the study's or instrument's consistency, predictability, dependability, stability, and/or accuracy and the establishment of reliability rests on replication (Guba and Lincoln, 1989). To assess reliability of the codebook, the two interviewers coded a sample transcript taken from interviews conducted by both interviewers. The coded data of each interviewer for the sample was compared by a third member of the evaluation team to identify the degree of similarity between the two coded sets of data. The two sets of coded data were concluded to be highly similar and the method was concluded to be reliable.

Independently from the interviewers, the transcripts were given to a panel of programme trainers who were asked to code the interview data into 'categories of meaning', in the same way as the two interviewers had done, in order to assess the validity of the data code-book. As the trainers had no prior knowledge of the code-book that has previously been developed, any differences between the code-books were identified and resolved.

5.3.8.4 Application

The stakeholder interviews were conducted initially in 1993 and again in 1995. In each case, the interviews lasted approximately 1 hour and were audio-taped. As there were interviewees from the UK and Germany, two interviewers were used; a bi-lingual German national for the German interviewees; and a UK national for UK interviewees. All interviews were conducted in English as English is the corporate language and all professional staff in Germany are fluent English communicators.

Interviewees were assured that the individual data would remain confidential. Despite these assurances however, several interviewees in Germany were reluctant to have the interview audio taped, and were therefore excluded from the process.

The 1995 interviews were conducted in the UK and Germany with 32 stakeholders (Table 5-XIV). They were intended to explore and extend many of the findings from the 1993 interviews. The same stakeholder interview schedule was used.

	Number of Interviewees		
Stakeholder Groups	UK	Germany	Total
Senior Managers	3	3	6
Trainers/ Internal Consultants	2	2	4
Trainees	3	2	5
Managers of Trainees	6	6	12
Course Administrators/ Co-ordinators	3	2	5
Total	17	15	32

Table 5-XIV: Stakeholder Groups (1995)

5.3.8.5 Data Analysis

To analyse the stakeholder interviews, content analysis was used. Dane (1990) defines content analysis as a research method used to make objective and systematic inferences about theoretically relevant messages. The basic methodology used involved careful (and painstaking) review of the transcripts of the interviews and highlighting (coding) relevant pieces of text using the code-book. Relevant statements were subsequently extracted from the transcript.

The data from the interviews was analysed initially within stakeholder groups to identify group constructions. From these constructions, decisions were made as to which constructions were pursued. The selected group constructions were then analysed across stakeholder groups with the aim of identifying the inter-group differences in constructions.

To further explore the generalised beliefs about the causes of successful or unsuccessful outcomes of the programme identified as part of the content analysis, an attributional analysis (Stratton et al, 1988) of the data was conducted.

Attributional analysis is a method for identifying, extracting and coding beliefs about causal relationships from qualitative interview material. Attributions are defined as "statements identifying a factor or factors that contributed to a given outcome" (Joseph et al, 1993).

Each attributional statement was extracted from the transcript and then coded. The coding system involved:

1. Identifying the 'Actor' and 'Target', where the 'Actor' is defined as the person or group causing something to happen, and the 'Target' is the person or group to whom something happens. Further coding sub-categorised the 'Actors' and 'Targets' considered to be particularly important to the evaluation. These sub-categories were: Self (speaker), Company (whether in the UK or in Germany), the programme (including training course, trainers, change programme), Employees (of the company, colleagues), Management (within the company), Other.

2. Distinguishing attributions according to whether they refer to a 'Positive/neutral' outcome or a 'Negative' outcome.

3. Identifying 'Actual' events or outcomes, defined as an event which has occurred or is on-going, from 'Potential' outcomes, defined as an event which may occur in future depending upon certain conditions being present.

5.4 Summary

The evaluation was undertaken in two phases; Phase 1- pilot evaluation; and Phase 2 - implementation evaluation, with each phase consisting of several components. The evaluation was undertaken by members of the programme development and training delivery teams, however the level 3 and level 4 components of the evaluation were undertaken using external resources.

In the next the chapter, the results of the evaluation are presented and in chapter 7 the evaluation approach, together with the results it yielded are analysed.

Chapter 6

Evaluation Study Results and Conclusions

6. Evaluation Study Results and Conclusions

> *Si nous ne trouvons pas des choses agréables, nous trouverons du moins des choses nouvelles*[1]
>
> *- Voltaire (1756)*

This chapter presents the results of the evaluation study and the conclusions which were drawn from the results for the purposes of the training programme. The impact and utility of the evaluation results and conclusions are the subject of Part III of this study and are therefore not considered here.

6.1 Pilot Study Results

As the pilot study evaluation was largely informal, no written reports were produced and retained for the two formative elements of the evaluation; classroom observation and spoken participant feedback. However, from my observations of these processes, the data was provided to the development team. This played a key role in determining changes to the programme.

6.1.1 Group Brainstorm and Discussion Session Feedback Data

The outcome of the feedback session was recorded and the following information was provided to the development mini-teams.

6.1.1.1 Three aspects of the programme which worked well

In response to (card and call-up) questions 'Identify the 3 aspects of the programme which worked well' and 'Why?', the following data was collected.

[1] If we do not find anything pleasant, at least we shall find something new.

Group 1		Group 2	
<u>What</u>	<u>Why</u>	<u>What</u>	<u>Why</u>
All Behavioural Skills	• Recognised Weakness • Relevant • Variation • Well presented • enjoyable • Virgin subject	*Team working*	• Plenty of practice • Good feedback
TOPS *-Decision making matrix* *-Problem definition*	• Relevant • Well presented • enjoyable • Practised Module	*Methodology*	• Engineers follow logical processes
CFE Module	• Well presented • enjoyable • Clear Road-map	*Concepts*	• Give perspective • Focused use of tool
Group 3		Group 4	
<u>What</u>	<u>Why</u>	<u>What</u>	<u>Why</u>
Overview *Ability to apply quality tools in an appropriate way*	• Well thought out • Well structured • Professional • Well presented • Good notes • Can see application • Useful concepts	*All Behavioural Skills*	• Behavioural Skills can be used every day • Team building and approach
Quality Engineering *Experimentation* *TOPS Methodology* *FMEA Systems Approach*	• Well thought out • Well structured • Professional • Well presented • Good notes • Can see application • Useful concepts	*Customer Focussed Engineering*	• QFD approach to original design
Team work *Facilitation skills* *Communication skills*	• Well thought out • Well structured • Professional • Well presented • Good notes • Can see application • Useful concepts	*TOPS*	• Decision making via TOPS
		FMEA	• System hierarchy approach
		Experimentation	

Figure 6-I: Pilot Group Feedback: Three aspects which Worked Well

The data given in Figure 6-I is from the 4 groups and has been taken directly from the metaplan session of each group. All of the groups restricted their feedback to the content of the programme, choosing not to identify structure or instructional design aspects. All four groups identified People skills as an aspect which worked well. This was encouraging to the development team as this was not an area of training traditionally favoured by technically oriented engineers. The reasons why the groups chose people skills are varied, 3 groups stated that the people skills were applicable or relevant to their work.

Two of the groups identified the conceptual overview aspect of the programme in that it provided a perspective for the application of the tools. The remainder of the feedback referred to the full range of the technical content of the programme.

6.1.1.2 Three aspects of the programme which could be improved

In response to (card and call-up) questions 'Identify the 3 aspects of the programme which could be improved' and 'Why?', the following data given in Figure 6-II was collected.

The feedback data is not restricted to just the content aspect of the programme. Three of the groups identified the instructional design as an area which could be improved, citing timing and the relationships between the trainers' presented materials and the notes they received.

The balance of the emphasis placed on certain technical content. The FMEA module featured in three of the four groups' feedback. The reasons for this centre on the instructional design of the module.

Group 1		Group 2	
What	How	*What*	How
FMEA	• For each module need road-map up front to see where we are and where we're going • Delivery - use of paper in notes	*Too Little:* • *Group Dynamics (too light weight)* • *DCP content* • *C to C (no detail)*	• More Coverage
Process Management	• For each module need road-map up front to see where we are and where we're going • Delivery - use of paper in notes	*Too Much:* • *Experimentation can get over complicated* • *Maths can get too heavy*	• Give it a trim
Timing	• Delivery - use of paper in notes • Keeping to a practical and workable agenda • ALL THE TIME!	*Off Target:* • *FMEA (forgetable)*	• More conceptual background
		Teaching Methods: • *Syndicate tasks - many too ambiguous / unachievable* • *Case studies relevant to motor industry* • *Notes which relate to slides; core and practitioner in same folder* • *Changing of examples too much (Exp. / QE)*	• Input from delegates • Could try harder!
Group 3		Group 4	
What	How	*What*	How
Case Studies	• Better quality equipment • Better continuity	*Improve:* • *Time management* • *Lack of task objectives* • *FMEA module*	• no suggestions given
Course Materials & Organisation	• Better notes management • Time management	*Elements Missing:* • *Behavioural skills - one to one feedback, stress management* • *Presentation* • *Concept to Customer*	• no suggestions given
Delivery	• Better focus on understanding	*Require Personal Review:* • *Experimentation - complex subject requires re-reading* • *Experimentation - further case study* • *Process Management - what do I need to do as a design engineer* • *SQC- right questions* • *QE - more information and experience on loss function, etc.*	• no suggestions given

Figure 6-II: Pilot Group Feedback: Three aspects which could be improved

6.1.2 Questionnaire Feedback Data

The following data was collected from the 16 pilot participants who completed and returned the questionnaires as part of the pilot study review session.

6.1.2.1 Overall balance of programme

Participants were asked to indicate how they perceived the balance of three aspects of the programme in terms of the behavioural / technical skills, lecture / syndicate work, and conceptual knowledge / applications training. The data collected is presented in the following bar charts.

6.1.2.1.1 Behavioural and technical skills

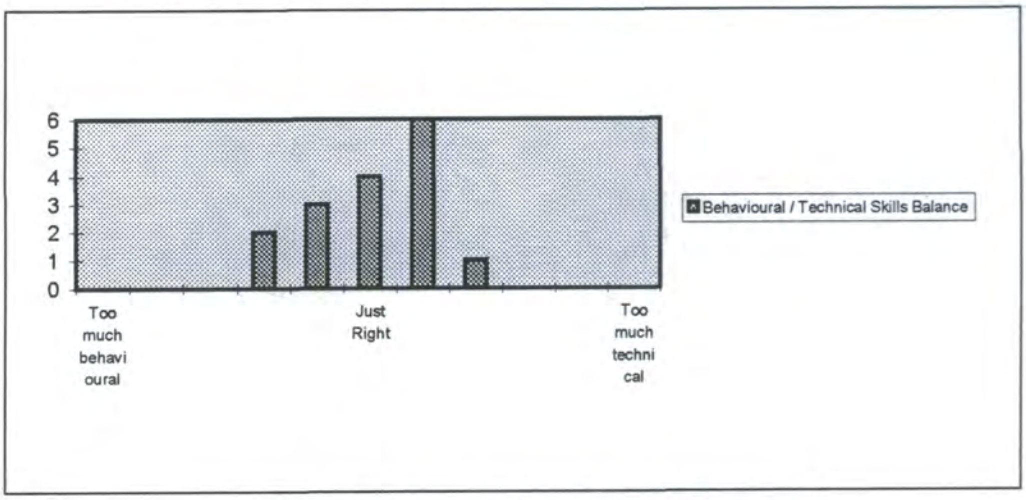

Figure 6-III: Balance of Behavioural and Technical skills

Referring to Figure 6-III, the bar chart indicates that overall the participants who responded perceived the balance to be right between the behavioural and technical skills, with 4 people indicating 'just right'. Five respondents perceived slightly 'too much behavioural' and seven slightly 'too much technical'.

6.1.2.1.2 Lecture and syndicate work

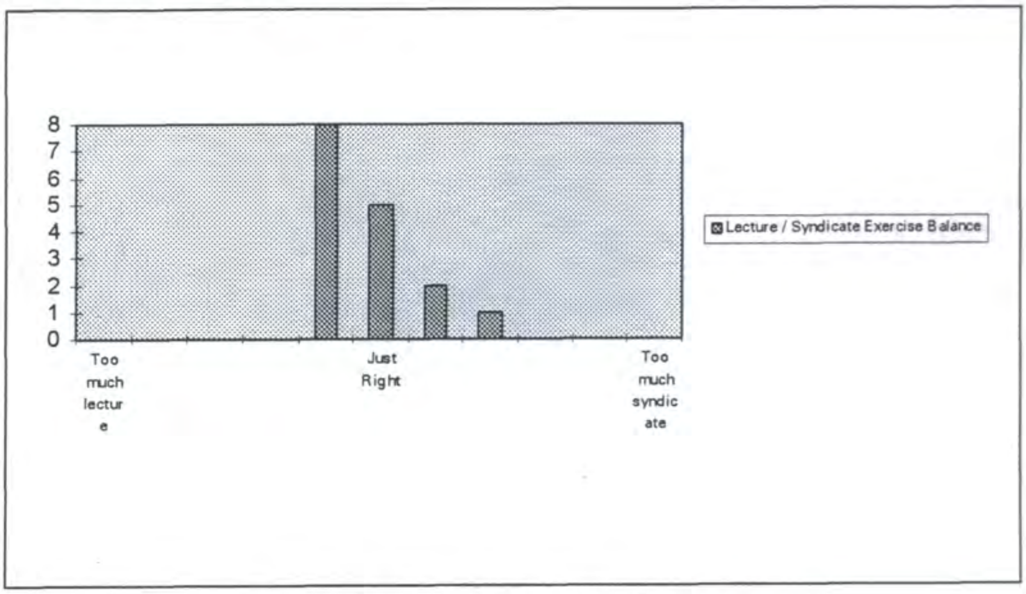

Figure 6-IV: Balance of Lecture and Syndicate Exercises

As part of the design process, the need had been recognised to find a balance between various styles of instruction used to deliver the programme. Lecture style instruction, which was prevalent within the company, was deemed to be ineffective in engaging students when used for long periods at a time. Whilst this approach had the advantage of presenting large amounts of information in a relatively short period of time, it was believed by the development team to be unlikely that students would be able to comprehend entire lectures in this way.

An instructional design guideline was established and applied as part of the design process, which was that all lecture style training sessions would be limited to 40 minutes duration. These would be interspersed with group based exercises, where the concepts which had been the subject of the proceeding lecture would be applied to a simulated engineering activity. In the main, this was achieved using the case study, but smaller stand alone exercises were also used.

The data collected for this question and presented in Figure 6-IV indicate that the pilot participants perceived a good degree of balance between the lecture style and exercise style of training delivery.

6.1.2.1.3 Conceptual knowledge and training

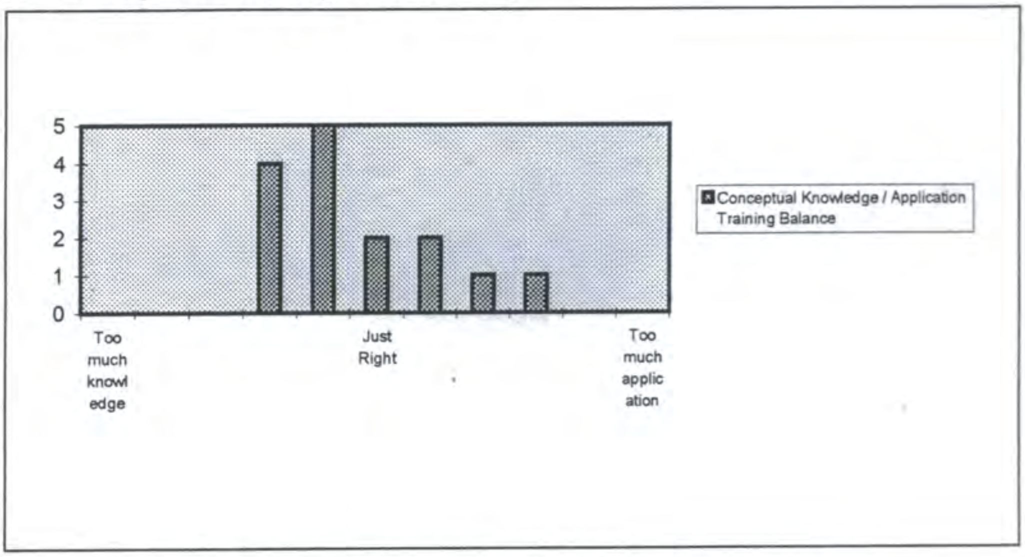

Figure 6-V: Balance of Conceptual Knowledge and Applications training

Given the complexity of the programme content and relationships between the content and to the engineering process, it was considered necessary by the design team that the programme should include training on the underlying principles and concepts of the quality methodologies, as well as the practical training of how the methods are applied to the engineering process.

Referring to Figure 6-V, the feedback from the pilot participants indicates that the balance between these two types of training was marginally weighted towards too much conceptual knowledge, however, this bias was not considered to be so significant that changes to the balance were required.

6.1.2.2 Short and long term contribution

Participants were asked to rate the short and long term contribution of the programme to professional development and ability to improve the quality of the company's products. A five point scale (1 = not important, 5 = very important) was used in terms of behavioural skills, technical (quality) skills, conceptual knowledge, application knowledge, training materials, and the case study. The data collected is presented below.

6.1.2.2.1 Behavioural skills

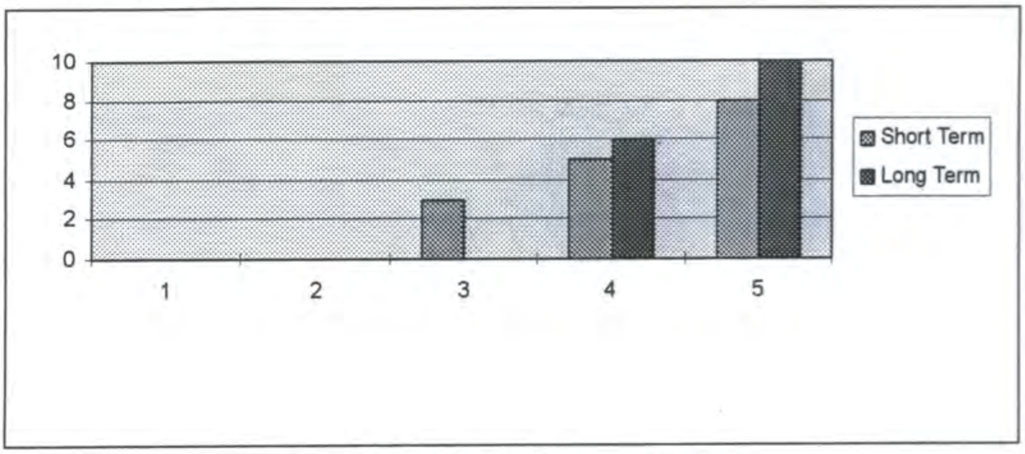

Figure 6-VI: Contribution of behavioural skills

The bar chart in Figure 6-VI indicates that high importance ratings were given for both the short and long term contribution of the people skills, with participants giving a marginally higher rating to the long term contribution.

As with the remainder of this series of questions, it is not possible to distinguish between the contribution to personal development and to the improvement of the quality of the company's products, without further investigation.

6.1.2.2.2 Technical quality skills

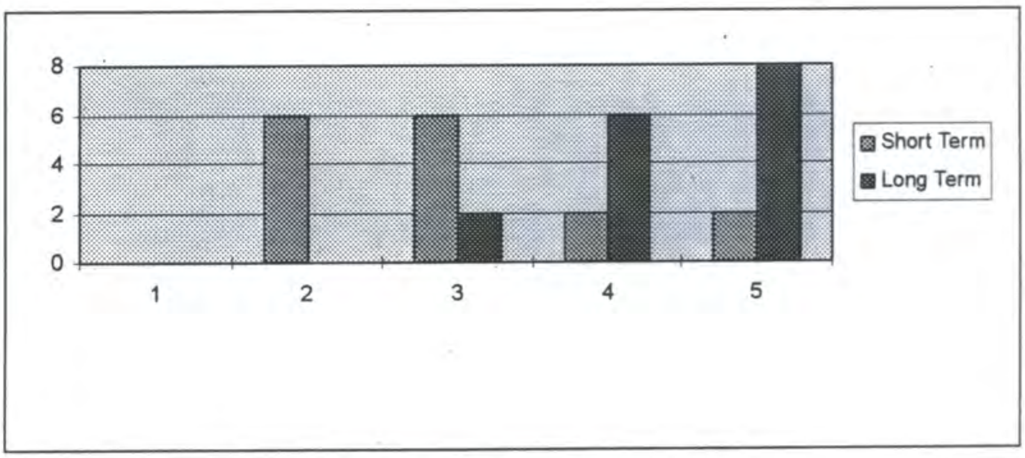

Figure 6-VII: Contribution of technical quality skills

The technical (quality) skills did not receive favourable importance ratings for their short term contribution, however they were recognised as being important to the long term contribution.

6.1.2.2.3 *Conceptual knowledge*

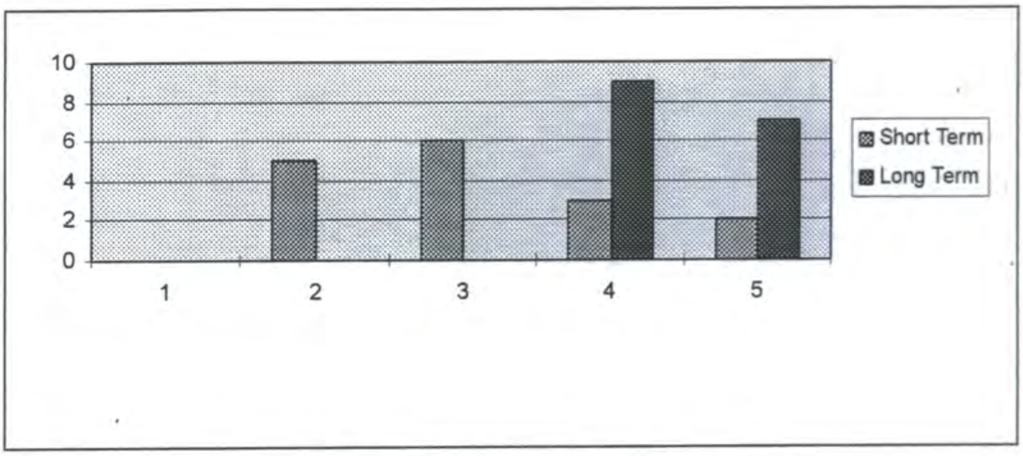

Figure 6-VIII: Contribution of conceptual knowledge

Conceptual knowledge was perceived to have limited short term importance, but high long term importance.

6.1.2.2.4 *Application knowledge*

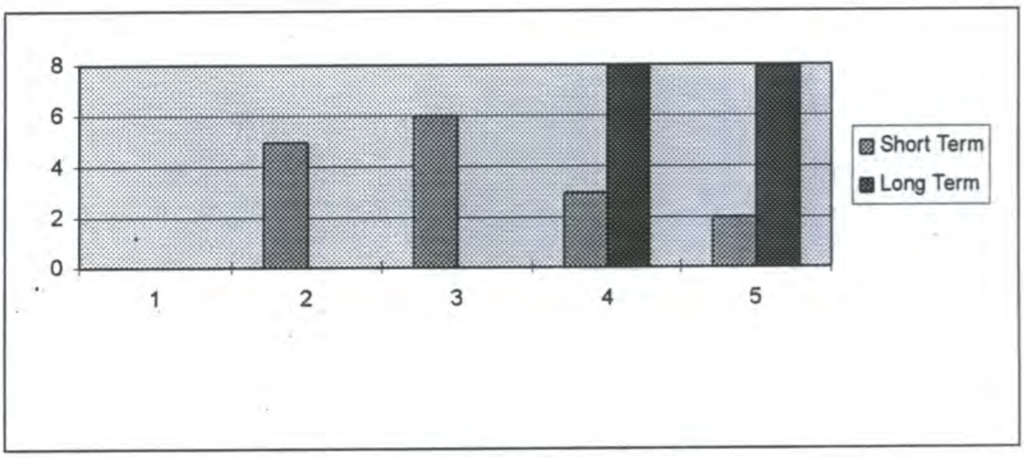

Figure 6-IX: Contribution of application knowledge

The feedback data for application knowledge mirrors that of conceptual knowledge in that participants perceived application knowledge to be more important in terms of its contribution in the long term, compared with the short term.

6.1.2.2.5 Training materials and Case Study

The perceived importance of the training materials and case study data is presented below in Figure 6-X and Figure 6-XI.

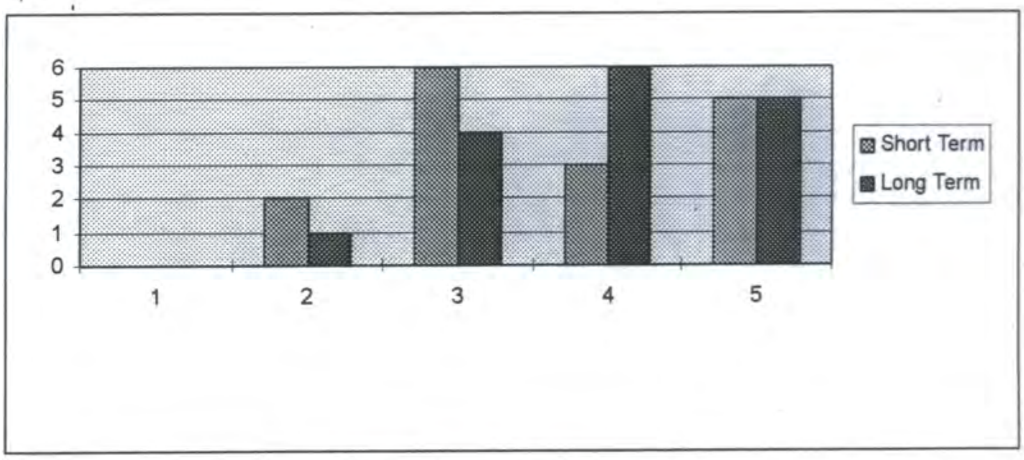

Figure 6-X: Contribution of training materials

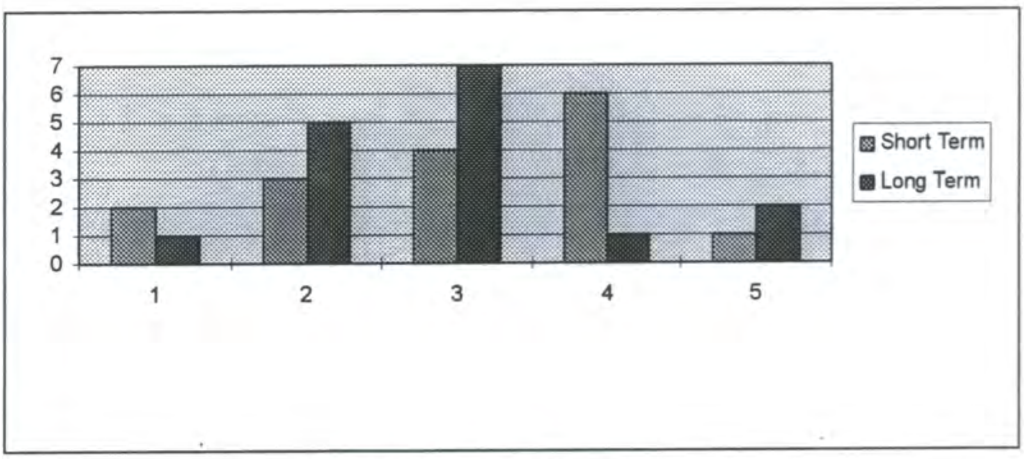

Figure 6-XI: Contribution of case study

The importance ratings for both the materials and the case study are dispersed across the rating range for both the short and long term contribution. Given the role of the materials is to facilitate teaching the content of the programme, this was not considered to be significant by the development team. The detailed feedback for each module provided the team with a better insight into the materials and case study. This feedback is presented and discussed in the following sections.

6.1.2.2.6 Module rating.

Participants were asked to rate each module in terms of 6 aspects: content, relevance, delivery (instruction), training materials, case study, and application to workplace.

A 5 point scale (1 = Bad, 5 = Good) was used.

The data collected is given in the tables below together with a bar chart of the combined data for each module.

6.1.2.2.7 Programme Foundation Module

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content		1		11	3
Relevance			1	6	8
Delivery		1	3	6	5
Materials			8	5	2
Case Study		1	7	4	3
Application		5	3	4	3

Table 6-I: Foundation Module data

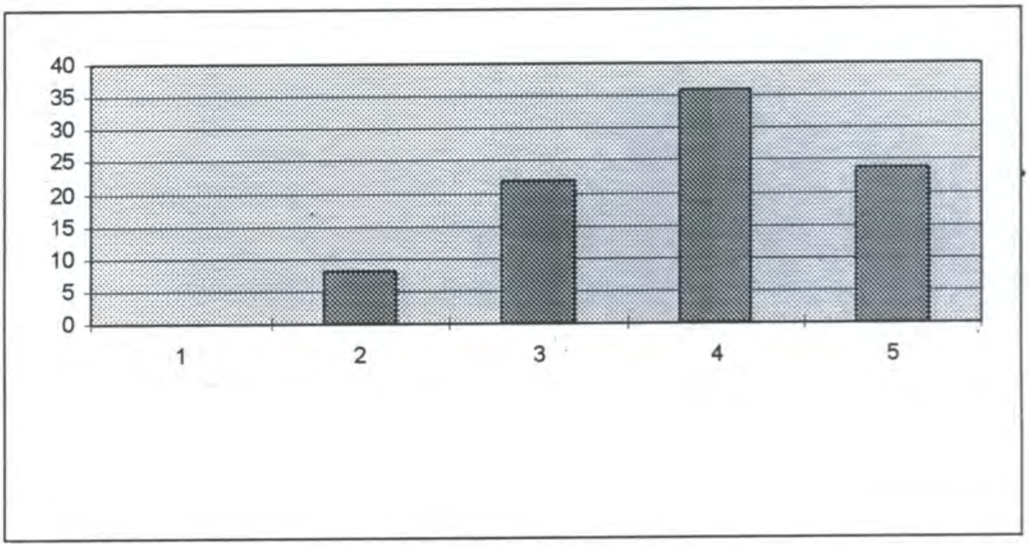

Figure 6-XII: Foundation module combined data

Referring to Figure 6-XII, the modal average rating for the module overall is 4. Referring to Table 6-I, the content and relevance aspects of the module received high ratings by respondents. The application aspect of the module received a comparatively low rating, but given that the objectives of the module are primarily concerned with the context of the quality methods and the content associated with these is considered at a conceptual level, these ratings were not considered, at the time, to be of concern to the development team. It was subsequently recognised however, as part of the evaluation of the implementation of the programme, that the people skills considered in the foundation module were perceived by participants to be part of the training, or instructional, process and not intended as a skill which should be applied by participants to their job. Although this was recognised at a later stage, earlier recognition would have indicated to the development team the need to place more emphasis on the people skills as being applicable to participants jobs. As the programme was still in its development stage at the time of the pilot, this would have been relatively easy to rectify. By not discovering this until full implementation, changes were more difficult and confounded by the complexity of the programme, by which time had been translated into four other languages.

The feedback given at the end of the pilot should have raised some concern, but the evaluation instrument used was insufficiently sensitive to yield this data, without further investigation.

6.1.2.2.8 Process Management Module

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content			5	5	5
Relevance		2	3	5	5
Delivery		4	2	7	2
Materials		1	7	5	2
Case Study		2	5	5	2
Application		4	3	4	4
Aspect	Level II				
	Bad 1	2	3	4	Good 5
Content			1	2	9
Relevance			2	2	8
Delivery		1	1	6	4
Materials		2	2	5	4
Case Study		1	6	3	2
Application		3	1	3	5

Table 6-II: Process Management Module data

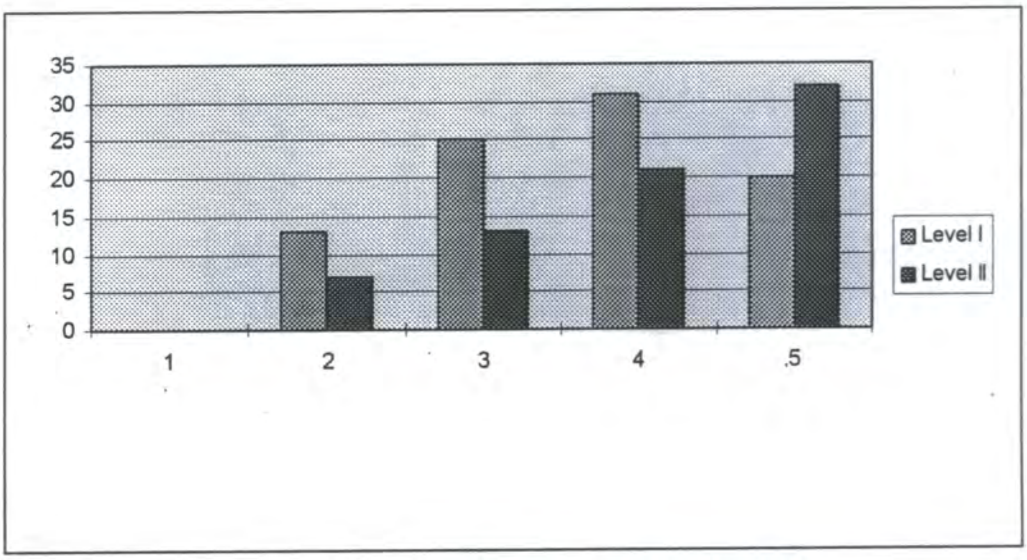

Figure 6-XIII: Process Management module combined data

Referring to Figure 6-XIII, the module was delivered at 2 levels and the data is presented accordingly. The level II course received more favourable feedback compared with the level I course. This difference is due largely to the feedback ratings given for content, delivery and materials aspects of the module (Table 6-II). From observations made during the delivery of the level I course, the flow and timing of the instruction was disjointed, with instructional items over-running the schedule.

6.1.2.2.9 Team Oriented Problem Solving Module

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content				3	13
Relevance				6	10
Delivery			1	2	13
Materials				5	11
Case Study				3	13
Application			3	5	8

Table 6-III: TOPS Module data

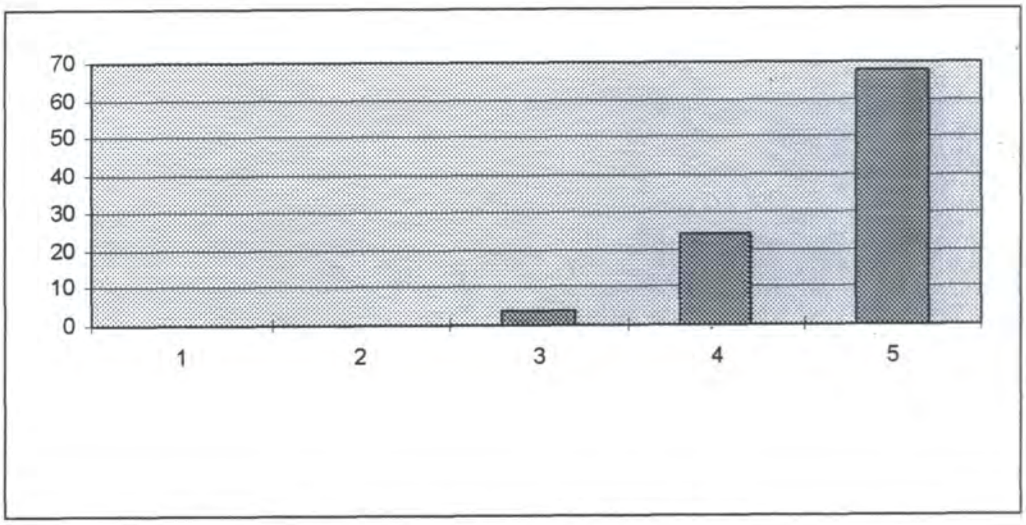

Figure 6-XIV: TOPS module combined data

The TOPS feedback is very positive (Figure 6-XIV) and this is reflected in Table 6-III, where all of the 6 aspects were rated high by respondents. The case study for TOPS received the highest ratings of all of the modules. The case study for the programme had been based on one which had originally been developed for TOPS prior to the advent of the programme. The scenarios presented as part of the programme had been taken directly from the pre-programme TOPS training course and were therefore highly developed.

6.1.2.2.10 *Failure Mode & Effects Analysis Module*

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content		2	3	8	3
Relevance			3	5	7
Delivery		4	4	3	4
Materials		2	3	5	5
Case Study		3	4	5	3
Application		1	2	4	8

Table 6-IV: FMEA Module data

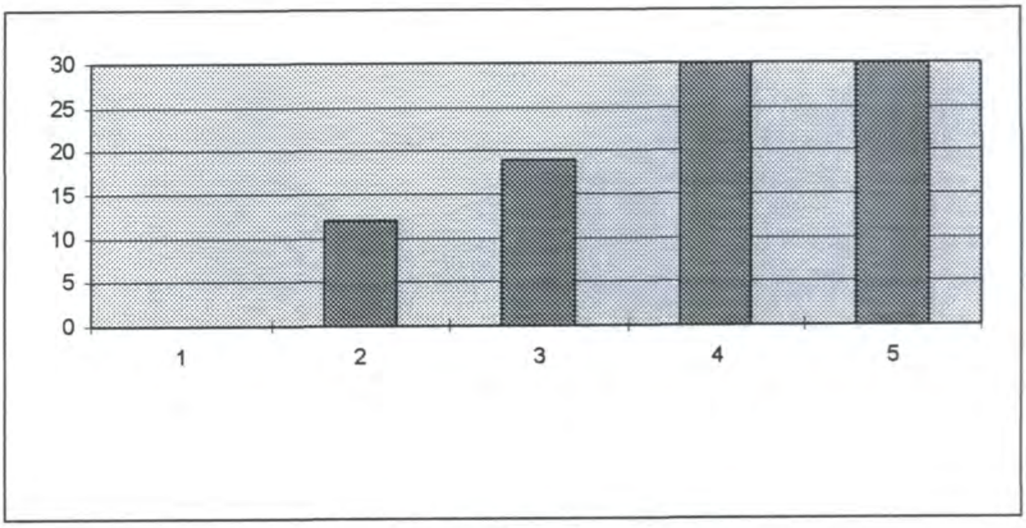

Figure 6-XV: FMEA Module combined data

Ratings given for the FMEA module presented in Table 6-IV identified the need for considerable improvement in the aspect of its content, delivery (delivery refers to the instructional processes), materials and case study. This feedback confirmed the observations made during the pilot delivery, which were similar to those made at Process Management.

6.1.2.2.11 Experimentation Module

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content			2	8	6
Relevance				7	9
Delivery				10	6
Materials			1	6	9
Case Study		1	1	7	7
Application		1	4	5	6
Aspect	Level II				
	Bad 1	2	3	4	Good 5
Content				7	8
Relevance				7	8
Delivery				5	10
Materials			2	4	9
Case Study			2	3	10
Application		1	4	1	9

Table 6-V: Experimentation Module data

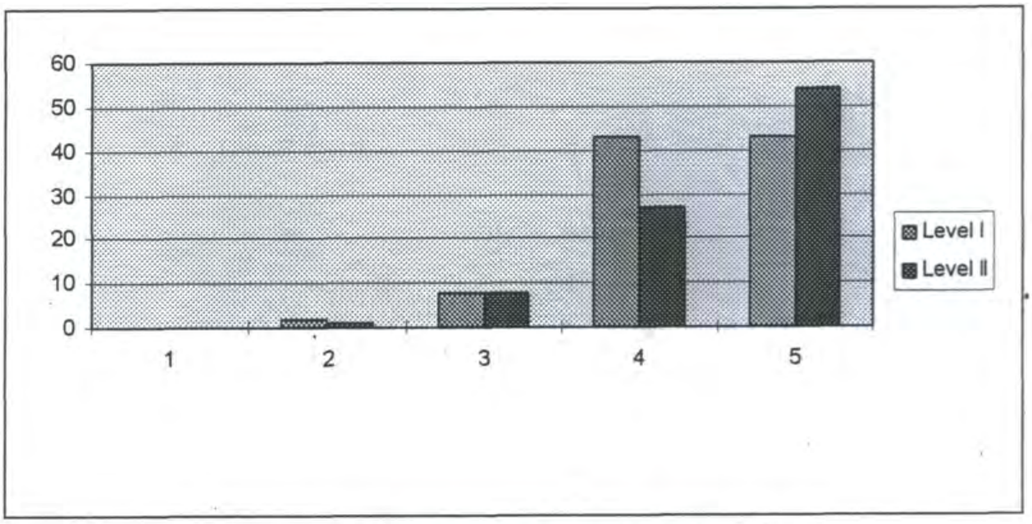

Figure 6-XVI: Experimentation module combined data

Referring to Table 6-IV, this module was delivered at 2 levels. For both level I and II, the module was given good ratings for content, relevance and delivery. Experimentation was relatively new to the company and although some training had been undertaken prior to the programme, it had been integrated with quality engineering under the label of 'Taguchi Methods'.

Within the programme, experimentation had been separated from quality engineering and the module started from the basic principles of experimenting with the factors of a product design or process as a means for planned improvement. The materials had been written as such and an ingenious simulation exercise was imported from Detroit, which involved a ball-bearing , a funnel and a ramp. The simulation enabled participants of the training to apply the concepts of experimentation to an actual situation - to maximise the time taken for the ball bearing to travel through the ramp and funnel assembly by experimenting with the factors which affect time taken. This was reflected in the relevance and delivery ratings.

6.1.2.2.12 Quality Engineering Module

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content			3	9	4
Relevance				9	7
Delivery			4	6	6
Materials		1	5	7	3
Case Study		2	1	9	3
Application			5	7	3
Aspect	Level II				
	Bad 1	2	3	4	Good 5
Content			1	8	6
Relevance				9	6
Delivery				9	6
Materials			4	5	6
Case Study		2	3	5	5
Application		1	4	8	2

Table 6-VI: Quality Engineering Module data

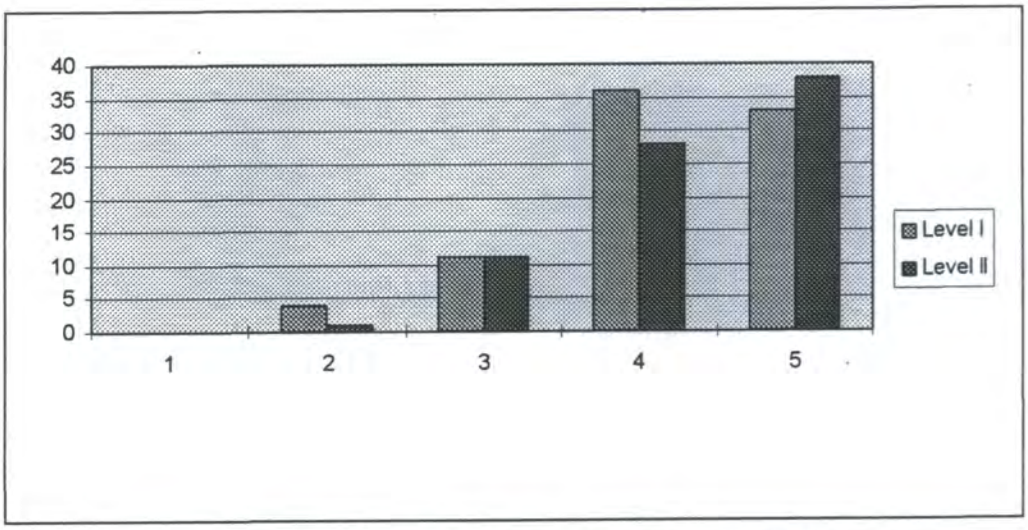

Figure 6-XVII: Quality Engineering module combined data

Referring to Table 6-VI and Figure 6-XVII, the Quality Engineering module received similar feedback to that of Experimentation. This was due mainly to its similarities of approach and delivery style.

6.1.2.2.13 Customer Focussed Engineering Module

Aspect	Level I				
	Bad 1	2	3	4	Good 5
Content				7	7
Relevance			1	6	7
Delivery			1	7	6
Materials			2	6	6
Case Study		1	3	7	3
Application		3	4	3	4
Aspect	**Level II**				
	Bad 1	2	3	4	Good 5
Content			1	5	7
Relevance			1	5	7
Delivery				7	6
Materials			2	6	5
Case Study			2	2	9
Application		1	5	3	4

Table 6-VII: Customer Focussed Engineering Module data

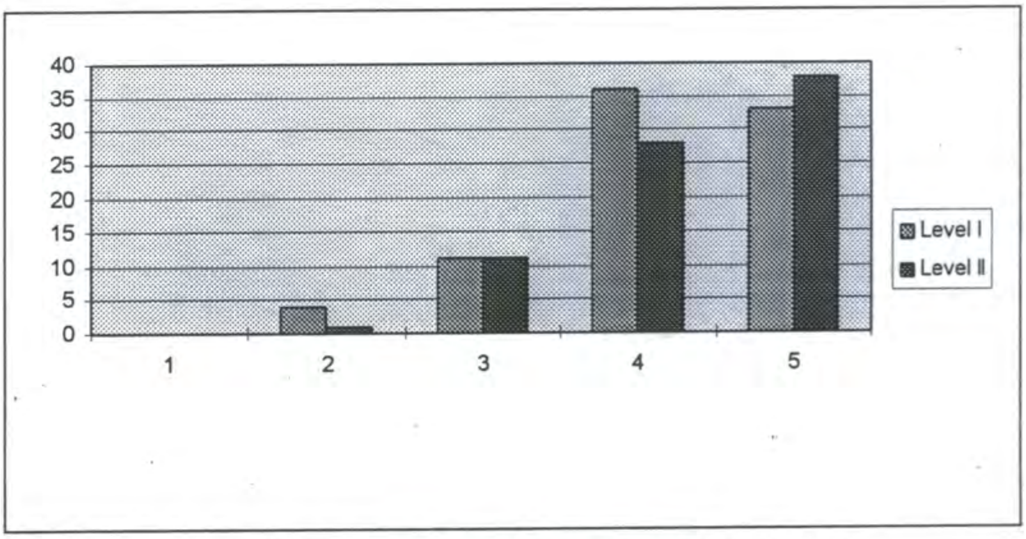

Figure 6-XVIII: Customer Focussed Engineering module combined data

Referring to Figure 6-XVIII, the CFE module was delivered at 2 levels. The ratings for the module are high and this is illustrated in Figure 6-VIII. Content, relevance, delivery and materials for both level I and level II received good ratings. Application of the methodology to the engineering process was not considered to be as good as the other aspects of the course, identifying the need to the mini-team, to offer a more convincing argument of the feasibility and benefits of using a customer focussed engineering approach as the engineering framework.

6.1.2.3 Overall Assessment.

As part of the overall assessment of the programme, participants were asked to identify: 1) content areas which they considered were missing, 2) content areas which they considered should be eliminated or reduced, 3) three specific aspects of the programme which improved their skills, 4) three areas of the programme which needed further improvement and 5) which parts of the programme did they believe important for their colleagues to attend.

The responses to questions 1), 2), 3), and 4) reflected the comments which had been made during the facilitated group feedback session. The responses to question 5) are given below.

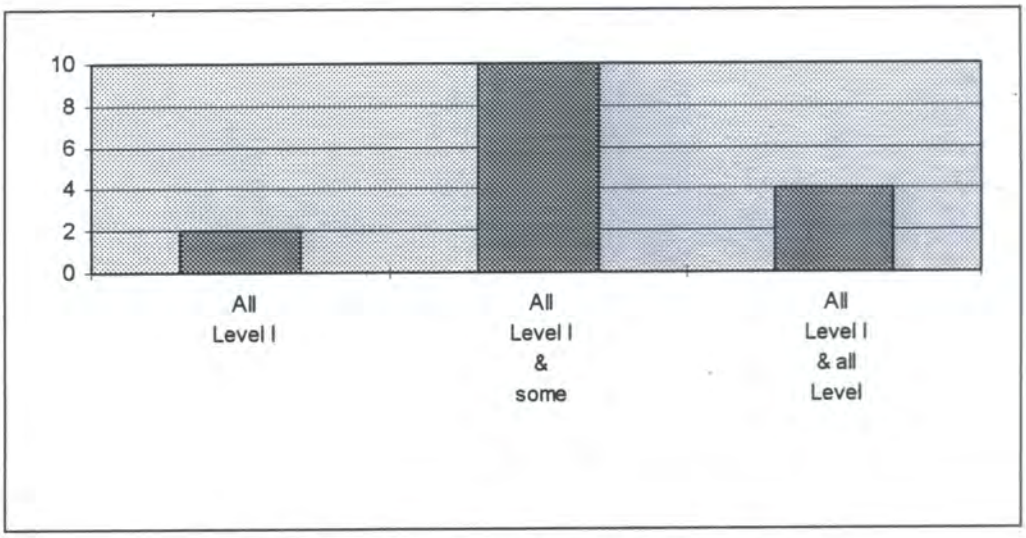

Figure 6-XIX: Parts of the programme considered important

6.1.3 Conclusions drawn from the pilot study evaluation

The overall conclusion of the development team was that the programme was likely to be successful as the basic design was sound. Specific areas of the programme had been identified as weak and required improvement:

6.1.3.1 Structural Improvements

The overall design structure of the programme was a key aspect to the development team as it set the parameters for the development of the individual modules. The programme was the company's first large scale in-house training initiative that was integrated and required participants to attend the whole programme in sequence. It was also the first attempt at combining a comprehensive range of technical and people skills for quality improvement into a single initiative.

6.1.3.2 Module Sequencing

The sequence of the modules had been based on the notion that the training should take engineers from where they were to where it was widely considered they should be. The majority of the company's engineers were, at that time, operating to a 'defect detection' paradigm, where most of their efforts to improve quality were employed downstream in the engineering process.

The widely held view amongst the quality Gurus, which was endorsed by the company's leadership, was that they should be operating to a 'customer driven and prevention' paradigm, where their efforts to improve quality would be upstream in the engineering process (Henshall, 1989). To this end, the programme sequence before the pilot consisted of an overview module (1. Programme Foundation module) followed by the down-steam modules (2. Process Management module and 3. Team Oriented Problem Solving module), followed by the up-steam modules (4. Failure Mode and Effects Analysis module, 5. Experimentation module, 6. Quality Engineering module and 7. Customer Focussed Engineering module).

The feedback from the pilot largely supported this sequence, although there was some discussion about the training following the chronological sequence of the engineering process (up-stream to down-stream), and it was decided that the sequencing of the modules aided learning, with the exception of the Process management and Team Oriented Problem Solving modules. As TOPS consisted of many of the introductory team concepts and as it was considered to be further down the engineering process than Process Management, TOPS was moved to become the second module in the sequence with Process Management as the third module.

6.1.3.3 Content and Balance

With few exceptions, the scope and depth of the content of the programme was considered to be appropriate for its intended purpose. The balance of the behavioural (people) and technical skills received favourable feedback, as did the balance between conceptual knowledge and applications training.

Also, and very significantly, the balance between lectures and syndicate exercises was favourably received by the pilot study participants. This represented a major shift from traditional in-house training. Lecture style instruction enables the transmission of a large amount of information relatively quickly. Whereas syndicate exercises tend to be very time consuming. It was considered to be important by the programme development team that a mix of lecture style training and syndicate exercise style of training was used if participants were to be fully engaged in the training. The feedback from the pilot study supported this view.

6.1.3.4 Programme Foundation Module

The pilot study Foundation module was delivered in 2 parts (2 days plus 3 days). From observations and end of day spoken feedback, it was concluded that the two parts would be delivered together. The advantage of making this initial module a residential course was muted at this stage. Subsequently, as a result of pressure to reduce the overall duration of the programme, the module became a two and a half day residential course utilising the evenings This meant that much of the content was unchanged as 28 hours of training was completed during the residential course.

The people skills content of the module (and of other modules) underwent some restructuring and development to better integrate the concepts with the technical skills.

6.1.3.5 Process Management module

Of all the topics within the programme, Process Management concepts had been the subject of the most training prior to the programme. Extensive Statistical Process Control training had been undertaken in the company since the early eighties. Process Management differed from earlier training in two main ways. 1) It was presented in the context of the overall engineering process and the technical and people skills necessary to support and 2) the emphasis on the programme was on the range of disciplines of process management and not just of SPC which had been the case previously.

In the planning stages leading up to the pilot study, the course developers had struggled with whether to include SPC starting from basic principles and risk going over previously covered material, or to assume that the participants coming to the module had a basic understanding of the concepts.

From observations and feedback from the pilot, it was concluded that the basic assumption that participants coming to the module would already have a fundamental understanding of SPC, and to reinforce this, the SPC Interactive Video was made a pre-requisite to course attendance. As the interactive video was available in all European (Company-location) languages and interactive video facilities were available at all Company locations, participants would have ample opportunity to gain the necessary understanding of SPC prior to attending the module.

It was also concluded that all materials required refinement and editing. This was true for all programme training materials.

6.1.3.6 Team Oriented Problem Solving module

Of all the module in the programme, the TOPS module was the most developed. It had been largely based on an existing training. This was reflected in the feedback that was received and although a few changes were made to the content, the course required little change.

6.1.3.7 Failure Mode and Effects Analysis module

Despite the emphasis placed on FMEA by the company in the years prior to the programme, the topic had not been the subject of any formal in-house training. The basic FMEA discipline is relatively simple, and yet few examples of good FMEAs could be found by the development team when writing the course. This view is shared by Johnson (1997).

Observations and feedback highlighted weaknesses in the instructional design of the module. From comparisons with other modules, where the topic methodology is structured into a staged framework, the mini-team concluded that a similar structure was required for FMEA. The technique was revisited and structured into a 6 stage framework (as discussed in chapter 4) and this framework was subsequently used as the foundation for the instructional design sequence. This resulted in significant changes to the module.

6.1.3.8 Experimentation module

Apart from editorial inconsistencies in the training materials, both the level I and II experimentation module received positive feedback. This was consistent with the observations made and feedback received during the delivery of the module. The main area of weakness identified was the application of the methodology to the engineering process. This was concluded to be due to a) the abstract nature of the training, and b) the paradigm shift from a problem solving technique to a quality planning technique.

The team decided to address these issue by including more actual examples of the application of the methodology to the engineering process.

6.1.3.9 Quality Engineering module

As previously stated, the Quality Engineering module received similar feedback to Experimentation. Similar instructional design changes were therefore made to the module, by inclusion of actual examples of the application of the philosophy to the engineering process.

6.1.3.10 Customer Focussed Engineering module

The case study and application aspects of the module were identified from the feedback to be require improvement. The case study formed an integral part of the instructional design and was therefore more complex than in other modules of the programme. The main area of concern was considered by the mini-team to be character and content inconsistencies in the case study. These were addressed accordingly.

The application aspect of the module suffered from a similar problem to that of experimentation and quality engineering. Customer Focussed Engineering represented a significant shift from the company's traditional approach to engineering and therefore actual case study examples were not available. Further, several previously failed attempts to successfully apply the methodology to the engineering process had done little for the reputation of the methodology. These failures were considered by the team to be due to applying the methodology to all aspects of the development process. A strategy of prioritisation of important vehicle characteristics was deemed necessary, and changes were made to the instructional design to emphasise this part of the QFD methodology.

6.1.3.11 The last 20% of improvement takes 80% of the effort

Measured in terms of both time and cost (FMC, 1995), approximately 20% of the programme development occurred before the pilot study was conducted. The remaining 80% came afterwards, despite the positive feedback that was received. This was largely due to two factors:

Firstly, by the time of the pilot study, the programme had become highly complex; it consisted of 36 days of training structured into seven modules, with four at two levels, giving eleven courses ranging in duration from 2 to 5 days. Linkages between the various concepts within and between the modules had been deliberately, and unintentional, designed into the programme, which meant that changes to one part of one module would set of a chain reaction through the programme.

Secondly, the translation of the programme led to many changes being made. The level of detail and understanding of the concepts involved in the translation process, identified material clarity issues, inconsistencies and errors. These led to reviews and changes to the source (English language) material. This was compounded by the continuous improvement and change of the source material, which meant that to support programme launch dates in German, Belgium, France, Spain and Portugal, some translation work was started before the source material had been finalised. The development team found themselves in a scenario where changes were being identified through using the material, from changes to other modules and from input to and from the translation process.

Regarding the amount of these changes which can be directly attributed to the evaluation of the programme, the evaluation highlighted aspects which required improvement. In the majority of cases, the feedback confirmed the opinions of members of the development team. Very few, if any, previously unidentified improvement opportunities were identified by the pilot study, however this does not invalidate or reduce the value of the feedback received. Whether the voice of the customer, in this case the participants in the programme, identify something new, or confirm what is already known, the act of working from the voice of the customer enables the producer (whether of cars or training programmes) to be customer focused.

6.2 Implementation Evaluation

The results and conclusions of the evaluation of the implementation of the programme are presented in terms of the components of each level of the evaluation (Table 6-VIII). An overall conclusion for the programme is drawn from these sets of results.

Level of Evaluation	Evaluation Components
Level 1 (Reactions)	Trainee verbal feedback process
Level 2 (Learning)	Knowledge assessment questionnaire
Level 3 (Behaviour)	Attitudinal questionnaire survey Behavioural observation
Level 4 (Results)	Stakeholder interview survey

Table 6-VIII: Implementation Evaluation Components

6.2.1 Reactions to the programme

As previously stated in chapter 5, participant reactions to the programme were not documented and cannot therefore be accurately reported here. However the use of spoken feedback as an evaluation method served a useful purpose to the programme trainers and from my own observations enabled them to continuously adjust their delivery style and other aspects of the training environment to meet the spoken needs of the learners.

6.2.2 Changes in Knowledge

Although pre and post knowledge questionnaires were developed for all modules of the programme, only the Programme Foundation module pre / post knowledge questionnaire set were developed into a reliable indicator of knowledge gain (see chapter 3). The remaining module knowledge change assessment questionnaires were not developed into reliable instruments and the causes and consequences of this are discussed and analysed in chapter 7. The following results, therefore, are for the Programme Foundation module only. Significantly, this module was intended to raise awareness of the quality techniques and underlying philosophy of the programme and in this sense a measure of knowledge gain is valid and appropriate.

The original pre/post knowledge tests which were extensively applied through to May 1995 were concluded to be unreliable and invalid (chapter 5: section 5.3.5.4). The following results are taken from a sample of six programme foundation modules from the second half of 1995 where the revised knowledge questionnaire was applied. A total of 81 engineers attended the six modules sampled.

The following table of results (Table 6-IX) reports the combined data for the six courses in terms of the percentage mean scores and range of all question items, technical content question items and people skills content question items.

	Pre		Post		Mean Change
	Mean %	Range (% pts)	Mean %	Range (% pts)	
Total	17.9	12.6	53.7	11.5	+ 35.8 %pts
Technical	15.4	11.8	52.0	12.5	+ 33.6 %pts
People Skills	25.7	23.7	60.0	17.3	+ 34.3 %pts

Table 6-IX: Programme Foundation Module Knowledge Test Results.

For both the technical and people skills content items, increases in participants levels of knowledge are indicated with 33.6 and 34.3 % points increase for technical and people skills respectively. From the range distributions for both pre and post tests and with respect to both technical and people skills, participant knowledge bands are consistent pre and post, however a reduction in range is evident for the people skills.

In addition to the pre and post course knowledge tests, a sample of 22 separate engineers completed a post course knowledge questionnaire 3 months after attending the programme foundation module. This sample had not previously been subject to the revised knowledge questionnaire. The results of the (+ 3 months) post knowledge test are given in Table 6-X.

	3 Months	
	Mean %	Range (% pts)
Total	56.2	16.8
Technical	52.39	17.7
People Skills	68.18	21.8

Table 6-X: Post 3 months Knowledge Test Results.

The results indicate that participants maintain the knowledge gain, with evidence of a further increase in their understanding of the people skills content of the Programme Foundation module. This may be due to participation in subsequent programme modules, however this data was not elicited from those sampled.

6.2.2.1 Changes in Knowledge Conclusions

It can be concluded from the results that engineers increase their knowledge of both technical and people skills by factors of 3.3 and 2.3 respectively as a result of participating in the programme foundation module. Furthermore from the limited sample studied after 3 months of attending the foundation module, the results suggest that the knowledge gain is maintained, however caution should be taken with this result as the knowledge maintenance may be due to participation in subsequent programme modules where many of the concepts of the programme foundation module are reinforced and elaborated upon.

The absence of knowledge gain data from modules other than the foundation module is a serious omission from this study. This was largely due to the non-reporting of results to the development and the emphasis which had been placed on the development of as reliable and valid module for the programme foundation module only. The reasons for this are examined in chapter 7.

6.2.3 Changes in Attitudes

Attitude change was measured in terms of programme participants' perceptions of the organisation and perceptions of quality. Three measures were developed (one for perception of the organisation, and two for the perception of quality) and applied to a group of engineers at three stages in the programme; 1) Prior to the Foundation module, where no training had been received, 2) After completing 3 training modules, and 3) After completing all level I modules.

The results from the application of these measures are presented below.

6.2.3.1 Perceptions of the Organisation

To recap from chapter 5, Perceptions of the organisation were defined in terms of four dimensions; principled management; security in help seeking; management efficiency; and trust in colleagues. The mean scores and range (expressed as one standard deviation from the mean) for the sample are given in Table 6-XI.

	Principled Management	Security in Help Seeking	Management Efficiency	Trust in Colleagues
Stage One Mean	47.891	72.803	58.041	74.503
Range	15.533	15.797	27.671	17.570
Stage Two Mean	46.895	71.487	62.309	75.163
Range	16.714	16.368	7.882	12.475
Stage Three Mean	43.403	67.569	63.333	70.185
Range	14.767	23.377	12.366	21.309

Table 6-XI: Perceptions of the Organisation Results

Interpreting the data, higher scores would indicate an increase (improvement) in the perceptions of the organisation. For all dimensions; Principled Management, Security in Help Seeking, Management Efficiency and Trust in Colleagues, there was no significant difference in mean scores at the three stages, with the range of scores remaining constant. This is better illustrated in Figure 6-XX.

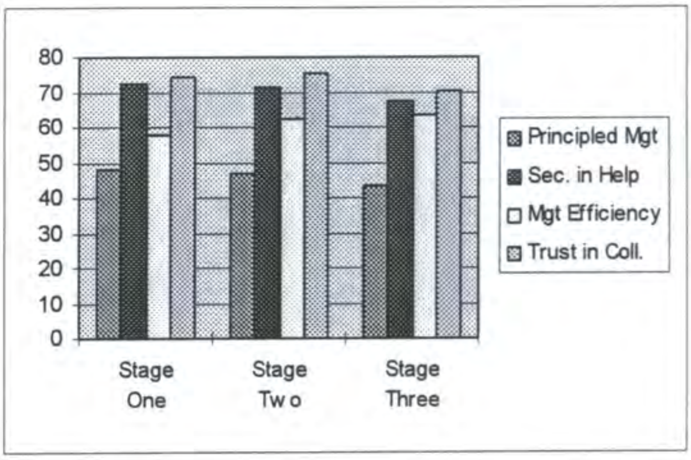

Figure 6-XX: Perceptions of the Organisation

The results of this limited study suggest that participation in the programme did not change participants' perception of the organisation with respect to the dimensions studied.

6.2.3.2 Perceptions of Quality

To recap from chapter 5, two measures were developed and applied to measure changes in participants' perception of quality. These were the Forced Choice Items Measure and the Likert Scale Items Measure.

6.2.3.3 Forced Choice Measure Results

For the forced choice measure, perception of quality was defined in terms of Innovation and Excitement. Table 6-XII gives the resultant score mean values and range (standard deviation).

	Innovation & Excitement
Stage One Mean	46.197
Range	27.691
Stage Two Mean	73.109
Range	31.494
Stage Three Mean	63.158
Range	25.339

Table 6-XII: Perceptions of Quality (Forced Choice Scale)

For perceptions of quality to improve, the mean scores using this scale are expected to increase. Examining Table 6-XII, the range of scores at the three stages remains constant, however there is a significant difference between the mean scores recorded for stage 1 compared to that recorded for stage 2 (Figure 6-XXI). Comparing the mean scores for stages 2 and 3, although a difference (decrease) is apparent, this is not considered to be significant.

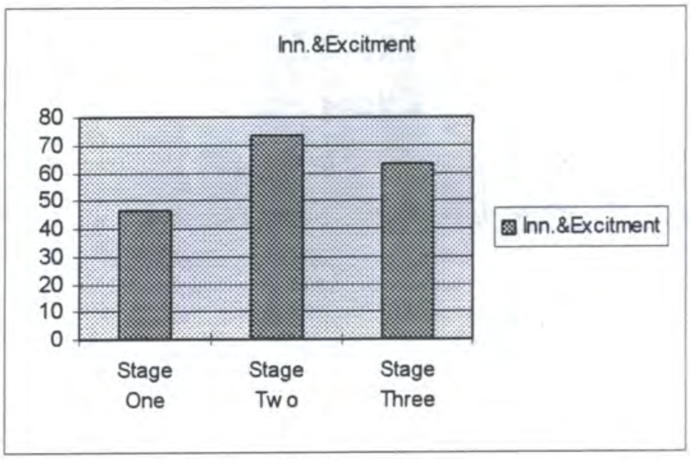

Figure 6-XXI: Perceptions of Quality (Forced Choice Scale)

These results indicate an improvement in participants' perception of quality, in terms of innovation and excitement on their introduction to the programme, but that no further improvements are made in this respect as they progress through the programme.

6.2.3.4 Likert Scale Measure Results

The likert Scale measure of perceptions of quality was defined in terms of four dimensions; customer empathy, creative engineering, product innovation, and open to new approaches. The results of the questionnaire are given in Table 6-XIII below.

	Customer Empathy	Creative Engineering	Product Innovation	Open to New Approaches
Stage One Mean	72.427	79.425	57.917	49.933
Range	12.086	13.182	20.101	16.056
Stage Two Mean	69.969	82.135	68.192	52.070
Range	10.018	14.944	14.467	18.633
Stage Three Mean	73.592	86.550	74.659	59.844
Range	11.496	12.046	17.731	19.316

Table 6-XIII: Perceptions of Quality (Likert Scale)

From Table 6-XIII, there is evidence of increases in mean scores for Creative Engineering, Product Innovation and Open to New Approaches as participants progress through the stages of the programme. Further analysis however, found the changes between stages 1 and 2 and between stages 2 and 3 not to be statistically significant. The difference between stages 1 and 3 using the same analysis was found to be significant. This is better illustrated in Figure 6-XXII.

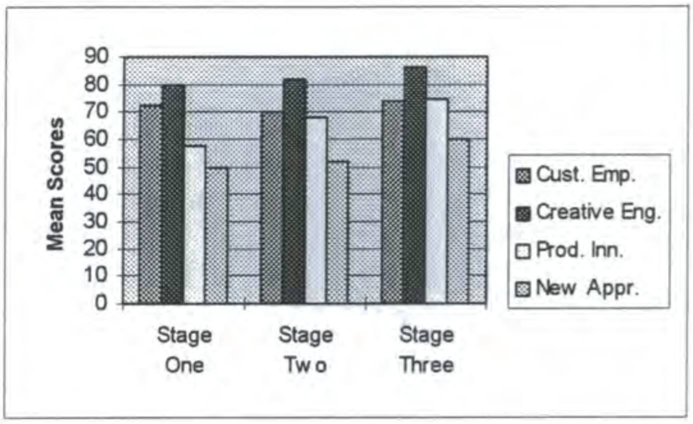

Figure 6-XXII: Perceptions of Quality (Likert Scale)

There is no evidence of change in the Customer Empathy aspect of participants' perception of Quality.

6.2.3.5 Conclusions

From the results of the perceptions of organisation data, it was concluded that the programme had no effect on participants. This was disappointing in that aspects of the programme were designed to shift participants' perceptions, however, aspects of the definition of perceptions of the organisation did not directly correlate with the programme objectives. Although this pointed towards a refinement of measurement instrument, no further development was made.

Regarding participants' perceptions of quality, the results indicated a (positive) shift in participants' perceptions. This correlated with the aims of the programme modules in that their perceptions shifted from a downstream and negative conceptualisation of quality (the subject of modules 2, 3, and 4; TOPS, Process Management, and FMEA respectively, where the emphasis is placed on puttying quality right) to upstream and positive conceptualisation of quality (the subject of modules 5, 6, and 7; Experimentation, Quality Engineering and Customer Focussed Engineering, where the emphasis is placed on defining quality which is exciting and performance driven by the customer).

The results also indicate the perceived importance of innovation through the quality spectrum (downstream to upstream), again conducive with the programme (people skills) aims.

Overall, the results for perceptions of quality indicated a very positive change in participants' attitudes resulting from the programme.

6.2.4 Changes in Behaviour

The following results are from observations made and analysed using the behaviour rating scales of two groups applying the FMEA methodology to the engineering process. Group A undergone training, whereas Group B had undergone no training.

The reader should note that the two groups were observed for their people skills behaviours only. The application of the FMEA methodology was not considered as part of this study.

Each Group was observed on 3 occasions and rated using the scales. For each set of observations, the group completed their own self evaluation using the scales.

The mean behaviour ratings for each of the people skills categories given in Table 6-XIV. The reader will recall from chapter 3, these ratings are established using a 7 point rating scale (1 = poor performance and 7 = excellent performance).

	Self-Evaluation Rating		Observer Rating	
	Group A	Group B	Group A	Group B
Warm-Up	3	2	2	2
Team Roles	4	3	3	2
Task/ Methodology	6	N/A	5	N/A
Process Review	3	2	1	1
Warm-Down	2	2	2	1
Listening	6	4	5	3
Questioning	6	4	5	3
Facilitation	5	3	4	2
Maintenance	5	3	4	2

Table 6-XIV: People Skills Behavioural Ratings

Analysis of Table 6-XIV reveals that the ratings for Group A were consistently higher for team roles, listening, questioning, facilitation and maintenance of the team. This was true for both the observer's ratings and the self evaluation ratings. The differences between observer ratings and self evaluation ratings suggest that group members overestimate the effectiveness of their team performance.

6.2.4.1 Changes in Behaviour Conclusions

From the application of behaviour rating scales to assess changes in behaviour with regard to elements of the programme's people skills, the following conclusions can be drawn:

recognising the limited scope of this application; to an experimental group (group A) and a control group (Group B), the feasibility of using the instrument as part of a continuous evaluation feedback process is untested. The implications of this are discussed in chapter 7. Despite this, the data suggests that the scales provided an assessment instrument which is reliable and valid.

With respect to the changes of behaviour, there is evidence that participants behaved as intended as a result of the training and that their behaviour is improved compared to engineers who have not participated in the programme.

From this limited evidence, it can be concluded that participation in the programme results in observable changes in behaviour.

6.2.5 Programme Effectiveness

To recap from chapter 5, the Programme's effectiveness was evaluated using a stakeholder interview process where interviews were conducted in 1993 and again in 1995 in both the UK and Germany. The stakeholder interview process employed a semi-structured interview technique which explored the areas of involvement; objectives and achievement, content, impact on customers / suppliers, barriers (to success), organisational climate, perceptions of the training department, and future vision (of the programme and the company).

The interview data was analysed using content and attributional analysis methods. The content analysis findings from the 1993 and 1995 interviews are given in the following pages.

6.2.5.1 Stakeholder Interviews: 1993

These data are from an initial round of interviews carried out in 1993. Overall, there was a high degree of congruence within stakeholder groups and there were some significant differences in perception between groups. However, the vast majority of differences in perception occurred cross culturally, between UK and Germany based stakeholders.

The findings are reported in terms of the major category headings.

6.2.5.1.1 Involvement

Predictably, the degree of involvement in the programme varied significantly between stakeholder groups. For those stakeholders for whom the programme constituted a major part of their job role (i.e. the Core Team and Trainers), interviewees described a high level of personal involvement in the success of the programme. In contrast, stakeholders who were essentially customers of the programme (i.e. Trainees and Managers) described their involvement in the programme as a part of their job. The majority of Managers interviewed had only participated in the Programme Foundation Module. Regarding future participation in the programme, there was a wide disparity between stakeholder groups, with many managers of trainees planning to attend one or two further modules of the programme, subject to their job commitments and releasability for training.

The majority of Managers and Trainees interviewed had attended the Foundation module following an invitation from the training department. Some had elected to attend further modules, but the majority did not have plans to take the initiative to schedule their own training.

With respect to module sequencing, few Managers or Trainees perceived the importance of attending the programme in sequence, regarding each module as a separate stand alone training course. Several of the Trainers expressed the view that each module should be named simply "Module 1 to 7" to discourage Managers and Trainees from pick and choosing the modules they wanted to attend on the basis of the module titles.

Regarding the programme's perceived impact on the future competitiveness of the company, all groups perceived the success of the programme to have a direct impact. However there was a wide disparity between different stakeholder groups in how much control they felt they had over the success of the programme. The predominant perception of the Managers and Trainers was that they had a high level of control, with many regarding themselves as programme "champions", whereas the Trainees and Course Designers felt that they had little control over whether the programme was successful.

6.2.5.1.2 Achievement/Objectives

There was consensus within and between stakeholder groups with regard to what the programme could achieve. Communication and teamwork across the organisation was regarded by all groups to improve as a result of the programme, which in turn would result in improvements in the quality of company's products and services. However, there were significant differences between Managers, Trainers and Trainees as to what they perceived as some of the outcomes of the training. Trainers and Trainees perceived "increased competitiveness" to be one of the most important outcomes of the programme, whereas the Managers spoke of this outcome less frequently.

Referring to Table 6-XV, cross national differences were observed in the interview data from the UK and Germany. Within the Trainer stakeholder group, German trainer-consultants referred explicitly more frequently to quality improvement (7 interviewees) and improved competitiveness

(7 interviewees) as outcomes, whereas the UK based trainer-consultants spoke more frequently of culture change (6 interviewees) and increased customer satisfaction (5 interviewees).

Perceived Outcomes	UK Based Trainer-Consultants (n=8)	Germany Based Trainer-Consultants (n=8)
Quality Improvements	6	7
Economic Benefits	4	4
Improved Competitiveness	4	7
Culture Change	6	3
Customer Satisfaction / Orientation	5	3

Table 6-XV: Trainer Perceived outcomes

Referring to Table 6-XVI, however the trainees' perceptions of the outcomes of the programme were consistent between UK and Germany, with the most important outcomes being quality improvement and culture change.

Perceived Outcomes	UK Trainees (n=8)	German Trainees (n=8)
Quality Improvements	6	6
Economic Benefits	3	4
Improved Competitiveness	4	4
Culture Change	5	5
Customer Satisfaction / Orientation	1	2

Table 6-XVI: Trainee Perceived outcomes

Within the Manager Stakeholder group, UK-based managers spoke more frequently of behaviour change and customer satisfaction as an outcome of the programme compared to their Germany based colleagues.

Differences between stakeholder groups was evident with regard to the relevance of the programme's technical and people skills to the engineering process. Trainees and Managers agreed that the technical content of the training was highly relevant but the people skills were of limited relevance. For application to their jobs, the people skills had to be interpreted to suit the work environment. Trainers and Course Designers, however, believed that both the technical content and people skills could be used directly back in the workplace. All stakeholder groups recognised the importance of teaching engineers people skills, particularly team building.

6.2.5.1.3 Content

Regarding the content of the programme, all stakeholder groups acknowledged that the appropriate engineering disciplines were included in the programme. Furthermore they perceived the balance of people skills / technical content to be correct, although there were differences in stakeholder perceptions of the technical / people skills content ratio. Within the Trainees and Managers of Trainees groups, some perceived the programme to be 90% technical skills and 10% behavioural skills, whereas others perceived the programme to consist of 40% technical skills and 60% behavioural skills.,

There were some differences between stakeholder groups with regard to content and method of delivery of the programme. In the UK, Managers perceived an imbalance between lecture style and syndicate work style delivery, expressing a preference for greater lecture style delivery. Some Managers Germany described the warm-up and warm-down group processes as being "exaggerated" and some Managers in the UK perceived some of the people skills to be "difficult to put into practice, back in the work place". In comparison, all Trainees (UK and Germany) were very positive about the style of delivery and enjoyed the group work.

The observation and feedback process employed to coach participants on their group work was seen as being very positive by Trainer, Trainee and Manager stakeholder groups in the UK and Germany.

Concerning the agenda timing of the programme modules, the majority of trainers interviewed regarded the timing to be good or acceptable to present and train the skills adequately. However, some Trainees and Managers made comments to the effect that the time allocated for the Process Management module was too short.

Among the Manager stakeholder group, perceptions of module agenda timing were more varied, with some regarding module durations to be too short, particularly for participants who were new to the concepts within the modules. Other managers considered the modules to be too long, particularly where participants had previously attended training on the concepts contained within the modules concerned (i.e. SPC course for Process Management, and team building for People Skills).

Regarding overall scheduling of the entire programme, Trainees and Managers acknowledged that although the time frame was long, this was necessary if the skills were to be transferred to the engineering process and therefore considered to be acceptable by most interviewees.

Concerning the training content as a whole, there was widespread acclaim from virtually all stakeholders, across all groups. The majority of those interviewed stated that they had thoroughly enjoyed participating in the modules and praised the "relaxed atmosphere" which was seen as "promoting the learning process".

With respect to participant administration, Managers and Trainees said that the programme wasn't sufficiently well marketed within the technical community and needed to be increased to a wider audience and sustained over a long period. Comments were also made to the effect that invitations to potential participants should detail the specific topics covered in the respective module. Many of the Managers and Trainees believed that it would reduce the possibility of the programme becoming "another flavour of the month". A large majority (>75%) of stakeholders believed that the programme would be "continuous" and was unlikely to become "another flavour of the month".

Note: The term 'flavour of the month' is corporate parlance which is used to refer to company initiatives which are short term and tend to be fashionable, or a pet project of a senior manager - often regarded as 'a good idea at the time' but soon fading away to be replaced by the next latest fad.

Concerning training facilities, UK based Trainees and Managers liked the use of "off-site" training delivery centres which were considered to "enhanced the relaxed atmosphere" of the programme. Trainees and Managers in Germany made comments to the effect that the training should be conducted outside the Company as it would "avoid distractions caused by problems in their daily work".

Note: In the UK, two off-site training facilities were used to deliver the programme. Whilst in the general locality of the majority of Product Development and Manufacturing sites, they were sufficiently isolated to take participants away from their normal work environment. In Germany, a single training centre was used which was located within the major manufacturing site.

With regard to participant demographics on the programme, all stakeholders considered it preferable where a cross-functional mix of participants attended the programme. Many viewed this as providing engineers with the opportunity to network with engineers from other departments, allowing them to gain some insight into the role and responsibilities of other engineers. Another advantage of cross-functional participation given by stakeholders was that it made the training "more interesting" in that it increased the range of inputs by participants. Trainees and Managers of Trainees in Germany also believed that the programme should include participants from functions and departments which were adjacent to mainstream engineering, such as production supervision, purchasing and finance analysts.

In response to the question of how the content of the programme related to the company's philosophy of Total Quality Excellence (TQE) in engineering, there was considerable difference in responses between Trainees and all other stakeholder groups. Disappointingly, a majority (>75%)

of Trainees were unable to describe the relationship between the programme and TQE and were unaware of how TQE related to the company's 'mission, values and guiding principles' (MVGP).

Despite this, all stakeholder groups believed the programme would change engineers' perception of quality, however, many Trainees in the UK consider this to result in significant changes in the engineering process. This viewpoint was not shared by either the Germany based Trainee stakeholder group or the other stakeholder groups.

6.2.5.1.4 Impact on Customer/ Supplier Relationships

All stakeholder groups considered the programme would affect the company's relationship with external suppliers as engineers and the engineering process would require improved incoming quality of products and services. This in turn would drive suppliers to increase their technical expertise.

UK based stakeholders believed suppliers should be involved in the programme as part of cross-functional, or vehicle programme teams. Although the Course Designers considered concepts of the programme to be relevant for customer facing activities, none of the other stakeholder groups expressed a view regarding their involvement in the programme.

6.2.5.1.5 Barriers

With regard to potential barriers to the programme's effectiveness (see Table 6-XVII), Trainer, Trainee and Manager stakeholder groups consistently identified reduced time for attending training, with several members of each stakeholder group referring to pressures from budget constraints and workload as a contributing factor to engineer release for training.

	Trainers		Trainees		Managers	
Perceived Barriers	UK (n=8)	D (n=8)	UK (n=8)	D (n=8)	UK (n=8)	D (n=8)
Time for Training	6	3	5	1	7	5
Time to apply skills	6	4	3	3	2	1
Resistance to Change	5	3	4	3	5	1
Critical Mass	4	1	6	1	2	2
Consultancy Support	4	3	2	1	3	1
Dept. budget constraints	4	1	1	2	3	1

Table 6-XVII: Barriers to the programme

Overall, greater importance was placed on barriers to the programme from the UK based stakeholder, with Managers and Trainers UK based stakeholders placing greater importance on all factors identified than their German counterparts. UK based trainee stakeholders placed greater importance on four of the six barriers identified compared to their German counterparts.

Employee resistance to change was perceived differently within Trainer and Manager stakeholder groups, with many UK based stakeholders associating the resistance with a perception that the company's management style was too autocratic for small groups of engineers to be able to change the way they worked. Several Trainer stakeholders believed that for changes to the engineering process to occur, a 'critical mass' was required. This is a colloquial term used to refer to where a significant numbers of engineers within departments participate in the programme to generate momentum for change. This relates to reduced time for training.

6.2.5.1.6 Organisational Climate

The organisational climate was consistently perceived to be one of low morale and insecurity relative to the past. Among the UK based stakeholders, however, there was a view that this situation would improve over time. The Trainers perceived the climate to be more positive than other stakeholder groups.

Among Germany based stakeholders, many referred to the external economic climate and the industry downturn as a contributing factor to the greater degree of insecurity. Many UK and Germany based stakeholders believed that the programme would have a positive impact on improving the climate, however some Germany based stakeholders regarded external economic situation as the determining factor of organisational climate.

6.2.5.1.7 Perception of Education and Training

The perception of the Education and Training department varied significantly between UK and Germany. In Germany, all stakeholder groups initially described training as being an important integral part of their daily work. However, they did not perceive this area of questioning to be relevant and so the German interviewer felt that this area of questioning was therefore redundant.

Among the UK based Managers, Trainees and Trainers the majority had a relatively negative perception of the Education and Training department and although they perceived the introduction of the programme to have improved the departments credibility, some Manager and Trainee stakeholders perceived the programme to be an Engineering department initiative, as opposed to a training department one.

All stakeholder groups perceived the content and delivery method to be of higher quality than other training intervention programmes. Many regarded this to be due to the credibility of the trainers, who were Engineers and so could speak with authority and offer examples from 'real life' situations. Many Trainers however, regarded themselves as becoming isolated from their engineering departments and believed this would reduce their credibility and career prospects over an extended period of time.

6.2.5.1.8 Future Vision

The majority of stakeholders believed the programme would continue. Within the Managers, Trainees, Administrators and Trainers stakeholder groups, many regarded continued senior management support as being essential in sustaining the programme. Managers, Trainees and Trainers perceived the programme as receiving the required financial support from senior management.

All stakeholders perceived that the technical skills in the programme were likely to be transferred to the engineering process. Regarding people skills, however, many stakeholders saw the need for a majority of engineers to have participated in the programme before they were transferred.

Regarding the programme's impact on the success of the company, Trainees and Trainers believed that the programme would have a strong impact. The Core Team and Managers believed that the programme would have some impact on the future competitiveness of the Company, but it was only one part in the process.

6.2.5.1.9 The Programme in a European Context

Within the Trainee stakeholder group, some were not aware that the programme was a European initiative, however all other stakeholder groups viewed this as being positive. Germany based Trainers felt they should have more influence on the development of the programme, particularly regarding the cultural adaptation of the instructional design and the requirement for sequential module attendance. Some believed that communication between the Trainers and the Core Team could be improved and that having a solely UK based Core Team was sometimes a problem.

The majority of the Managers and the Trainees believed that the programme would improve communication between UK and Germany based engineers by providing a common understanding of quality improvement. Many stakeholders saw a need to extend the programme to the company's operations in Southern Europe. (Note: At the time of the interviews, programme deployment to France, Portugal and Spain was at an early planning stage).

6.2.5.2 Quotations from the Stakeholder Interviews

The following quotations (Table 6-XVIII) are extracted from interview transcripts with the programme staff; (i) the development team; (ii) the trainer-consultants; and (iii) the programme administration staff. These were considered by members of the evaluation group to give further insight and texture to the stakeholder feedback.

Stakeholder Group	Significant Quotation
Development Team:	*"If you don't change nobody will - you are the Company"* *"the programme has been seen as a quality skills programme, and accepted as such - I don't think people have actually really understood what we're trying to do on the people skills side"* *"As an individual I don't think I have a lot of influence, I believe that influence comes from the core team and the consultants together, ... it's over and above what we do on the training"* *"One of the things that happens at the Company is that people will say "well we're not going to do that because we don't see our managers doing it - when they do it, we'll do it." People don't see that they can do it - they are the Company and they can influence the Company"*
Trainer - Consultants:	*(Referring to the programme planning and implementation) "If the methods of the programme had been applied to the programme, some things could have been avoided"* *"Often we hear 'it is amazing that we have achieved so much in a short space of time'. The people are impressed about the success of the team, and the experience is more valuable than a three day theoretical seminar"* *"Personally, I am very convinced of the programme, and to my mind, it can take us to where we have to go. Regarding quality, with the methods we have been using in the past ten years, we did not reach it. In order to take that huge leap we need these methods especially when we see the Japanese vanishing into the distance"* *"We practise team work. The people learn that you have to consider other people, and take into account all viewpoints and that the result is marked by higher security and more information and influence"*
Administrators:	*"the programme (as a training programme) is of a higher level sophistication than anything else in the Company"* *(Referring to the programme implementation strategy) "We should be applying the principles of process management to the programme"* *"the programme will lead us to better quality .... and greater customer satisfaction .... it's absolutely vital"* *"It's the responsibility of every line manager to ensure that they are sending their people on the programme"*

Table 6-XVIII: Programme Staff Quotations (1993)

In contrast, the following quotations (Table 6-XIX and Table 6-XX) are extracted from interview transcripts with the customers of the programme; (i) the trainees; (ii) managers of trainees; and (iii) senior managers.

Stakeholder Group	Significant Quotation
Trainees:	*"Since the group was mixed so interestingly, it was a key event for me"* *"If the programme is continuous over years, at least I think very highly of the programme, because realistically, it is very good"* *"To transfer the skills , greater commitment is needed from the Supervisors"* *"I didn't even know that the programme was European"* *"That's a point, the trainers. They are one of us (engineers) so they know what they are talking about. That's a real advantage .... it also makes the course more efficient"* *"I was really enthusiastic about the Foundation module. Although I knew several topics already, there was a lot that was completely new, where I gained insight....I think the programme is absolutely necessary and in the present environment, essential for the survival of the Company"* *"...... the Taguchi experimentation. It is difficult to follow the techniques strictly, you put them aside sometimes. The problems arise because not all the people have the same level of knowledge and there are influences from outside which can stop you from carrying it through rigorously. An example; we had planned an experiment following Taguchi. What we wanted to obtain we did not achieve at the beginning. Thus we just put it aside; the approach was good, but we got into severe practical difficulties"*
Managers of Trainees:	*"....something they should try to do in the programme is to bring the different places, nationalities together and hope to break down the barriers and develop teamwork"* *"Even if the programme as such dies, the Company will continue living, the people will find ways of working sensibly"* *"By training engineers only (not Production Foremen/Group Leaders) we opened the classical interface, the engineers with the tie and the foremen with the overall. I think this is very bad. Leaving those who understand what it's about and those who don't"* *"Time, workload and release for training are major barriers to the programme being successful"* *(Referring to People Skills) "When one gets to the bottom of problems, then it is nearly always communication breakdowns. I believe it's important to include them. In fact, it's essential. If we improve our communication, then we can gain alot."* *"Ideally, in every department there would be one person who is trained in the same way as the trainers; if possible a young person that I could promote"* *"I'm really enthusiastic about the programme. I've been in the Company a long time....and I wish we had something like it ten years ago"* *"It's like everything, if you want it to be successful you've got to keep advertising, you can't just advertise once"* *"Top managers have been on the course but I don't see any evidence that they're using the stuff they were taught in meetings"*

Table 6-XIX: Programme Customer Quotations (1993)

Stakeholder Group	Significant Quotation
Senior Managers:	*"For change to take place .... far more resource has to be put in the learning and education process .... previously far too much emphasis has been placed downstream"* *"There is too much de-stabilisation. We need to freeze and stabilise. We're not providing the climate for these changes to mature"* *"Universities should train engineers in some of these skills, but they don't"* *(Referring to the programme) "It's not optional, it's a mandatory thing .... it has to be continuous and updated so that it stays the 'state of the art'"* *(Referring to the Education and Training dept.) "It's a shame that we rarely praise each other .... I can't think of many companies that are pushing ahead with so many education programs"*

Table 6-XX: Programme Customer (Senior Managers') Quotations (1993)

6.2.5.3 Conclusions and Recommendations from 1993

From the data provided by the 1993 interviews, conclusions were drawn with respect to; (i) programme outcomes; (ii) programme content and structure; and (iii) programme administration, with the intention of identifying improvement opportunities.

(i) Programme Outcomes:

The outcomes of the programme were perceived to be positive in terms of increased competitiveness and quality improvement for the company. This perception was particularly evident in the Trainers and Trainees stakeholder groups (UK and Germany), however UK based Trainers and Trainees perceived organisational behaviour to be largely "autocratic" as opposed to "team focused". This behaviour was regarded by many to be a significant barrier to the changes intended by the programme. Managers also identified economic and resource issues as programme barriers, affecting the release of trainees and the application of new learning.

In the UK, particularly, it was thought that the idea of putting a critical mass of trainees through the training programme would produce the necessary cultural change. A critical mass was an analogy used by a senior engineering manager to describe a condition where the majority of the company's engineering community had participated in some or all of the programme. As to how many people and how much of the programme constituted a critical mass was undefined, however within the training community it was interpreted as 60% of all levels (i.e. engineers; supervisors; managers) having attended relevant programme modules. In the case of engineers this was the level 1 modules which related directly to their role in the engineering process and for supervisors and managers this represented attending the Programme Foundation module.

Across all stakeholder groups, there was a desire for the programme to be successful and it was seen as an important programme. The feedback from the stakeholder interviews largely reflected participants' reactions to the training on completion of the modules.

(ii) Programme Content and Structure:

With regard to the programme content, the balance of people and technical skills was widely regarded as being correct. UK-based managers believed that there was too much 'group work' and would have preferred a more 'classroom style' of training. UK-based managers also described some of the people skills as being "difficult to put into practice, back in the work place".

Regarding training attendance, the modular structure and recommended attendance sequence and frequency, were perceived by managers (in the UK and Germany) as being difficult to meet. This related to resource constraints inhibiting the release of engineering staff to attend training.

The stakeholder interviews largely endorsed the content and structure of the programme.

(iii) Programme Administration:

With respect to awareness of the programme's existence and aims, it appeared that most knew what the programme was designed to achieve, however many stakeholder believed that the advertising of programme should be intensified. No conclusions were drawn from the data with respect other aspects of programme administration.

From these conclusions, the evaluation group identified four recommended actions (Table 6-XXI) which were reported to the training department management at the Programme Steering Committee.

Action #	Recommended Actions
1.	The programme should be continuous and delivered to other functions within the Company. It is also recommended that different levels of the organisation (e.g. Group Leaders) attend the Foundation module, so that the programme is not just seen as a discrete training programme, but as an organisational change programme.
2.	The programme should be 'advertised' more widely throughout the Company to ensure that engineers fully understand the objectives of the programme and the intended organisational change. In this way, the relationship between the programme, other Company initiatives and future corporate objectives can be better understood.
3.	The technical skills are more easily transferable back in the workplace than the behavioural skills and so this area could be investigated further, thus giving more detailed advice as to how to apply the behavioural skills is given in the training programme.
4.	The principles taught in the programme could also be linked to the selection and appraisal system, therefore increasing the likelihood of transfer of skills and behaviours learnt back to the work place. One example is the development of the Commitment to Quality psychometric test by the Nottingham Team.

Table 6-XXI: Recommended Actions (1993)

The recommendations largely reflected the views of the programme management. Actions associated with recommendations 1 and 2 had previously been initiated on the basis of trainer feedback to the programme management. With regard to recommendations 3 and 4, no specific actions resulted from the report.

The consequences of the 1993 evaluation are reviewed in Chapter 7.

6.2.5.4 Attribution Analysis Results from 1993

From chapter 5, the reader will recall that to further explore the generalised beliefs identified as part of the previous content analysis, attributional analysis of the data was undertaken. The results of the analysis are presented in the following sections.

From the interview transcripts, 1230 attributions were identified and coded. The attributions were analysed within and between stakeholder groups.

	Positive Attributions	Negative Attributions
Actual related attributions	319	448
Potential related attributions	298	165

Table 6-XXII: All Attributions (1993)

Considering all attributions identified, they can be analysed in terms of Positive/Negative and Future/Actual. Positive attributions are defined as statements or expressions which are favourable to the programme. Negative attributions are defined as statements or expressions made by stakeholders which are unfavourable to the programme. Future attributions are defined as statements or expressions made by stakeholders in a future context. Actual attributions are defined as statements or expressions made by stakeholders referring to factual events or phenomena. This high level analysis is given in Table 6-XXII, where the frequency counts for attributions coded as either Positive/Negative and Future/Actual are shown.

The table indicates that although interviewees described a large number of negative events occurring within the organisation, when referring to the future, they use positive statements to describe the outlook for the organisation.

Referring to the programme as an actor (the reader will recall from chapter 5 that the coding identified causal relationships between actors [causes] and targets [affected]), 413 attributions were identified, of which 277 were coded as positive and 136 as negative. The attributions were further coded in terms of actual and potential. Where actual referred to events which had occurred and potential referred to future expected events. The coding analysis is given in Table 6-XXIII.

	Positive Attributions	Negative Attributions
Actual related attributions	135	85
Potential related attributions	142	51

Table 6-XXIII: Programme as Actor (1993)

From Table 6-XXIII, the analysis of the positive/negative frequency ratio suggests that interviewees perceived the future effect of the programme more positively than they did at the time of interview.

Referring to the Company as an actor,212 attributions were identified, of which 76 were coded as positive and 136 as negative. Again, the attributions were also coded in terms of actual and potential. The coding analysis is given in Table 6-XXIV.

	Positive Attributions	Negative Attributions
Actual related attributions	37	107
Potential related attributions	39	29

Table 6-XXIV: Company as Actor (1993)

The analysis suggests that interviewees perceived the Company to have caused a significantly large number of negative events in the organisation. With regard to the future, interviewees perceived a more positive outlook, however this was not as favourable as for the programme as an actor.

6.2.5.5 Stakeholder Interviews: 1995

These data are from the 1995 follow-up interviews. As for the 1993 interviews, the findings are reported in terms of the major category headings.

6.2.5.5.1 Achievement/Objectives

All stakeholder groups perceived that the application of the technical skills would lead to quality improvements in the Company's products and services. However, few interviewees described the potential benefits that could be brought about by the people skills. Only the senior managers made reference to the importance of the people skills.

Without exception, stakeholders referred to potential benefits of the programme, with none offering actual examples where the programme had resulted in tangible improvements.

No UK / Germany differences in perceptions of achievements or objectives were observed.

6.2.5.5.2 Content

The content of the programme received widespread acclaim from virtually all stakeholders, across all groups. The majority of stakeholders described their participation in the programme as enjoyable and considered atmosphere created during the programme's training modules to be relaxed and conducive to learning.

Within the Trainee stakeholder groups, there were some differences between the UK and Germany based interviewees, with some the German interviewees describing the programme content and delivery as having too much "group work" and syndicate exercises. Preference was for lecture style delivery was expressed. Again, as in 1993, comments were made by the UK based Manager stakeholders which described the warm-up and warm-down group processes as being "exaggerated".

The people skills content was perceived by many Germany based Trainee and Manager stakeholders as being difficult to put into practice in the work place. Many did comment that they had found the people skills training to be enjoyable.

With regard to the duration of the programme, the majority of interviewees (>60%) from both the Trainee and Manager stakeholder groups perceived the whole programme to be too long. Many Trainees regarded this as a potential inhibitor to attending all modules.

As in 1993, all stakeholders perceived the programme to be of higher quality than other training programmes in terms of content and delivery style. Again, this was considered to be due to the involvement of the company's engineers as trainer-consultants.

Similarly to 1993, all interviewees believed that there had been little advertisement of the programme and that invitations to attend should detail the topics covered. The administrators / co-ordinators saw a need for the sequencing requirement of the programme to made explicit in joining instructions and marketing literature to ensure that all future participants were made aware of the requirement.

The majority of stakeholders believed that the programme would be continuous, and as in 1993, many felt this was dependent on continued senior management support. Senior managers recognised the need for a global quality programme.

Regarding the mix of participants on the training, all stakeholders expressed a preference for cross-functional teams.

Trainee stakeholders were unable to describe the relationship between the programme and TQE and their relationship to the company's 'mission, values and guiding principles'. Members of the Trainer group also made comments to this effect about the participants on the programme.

6.2.5.5.3 Impact on Customer/ Supplier Relationships

The involvement of suppliers in the programme, particularly as part of cross functional teams was mentioned by members of all stakeholder groups. This was perceived to be beneficial in that it would give them a common understanding of the company's quality methods as part of the engineering process.

6.2.5.5.4 Barriers

With increasing frequency, compared to 1993, Trainee and Manager Stakeholders considered the programme's inflexibility with regard to module attendance sequencing to be one of the main barriers engineers gaining access to modules 2 - 7. Many also perceived there was insufficient number of Programme Foundation modules scheduled to meet demands. Within the Trainee stakeholder group, many said they found difficulty with planning their training due to problems in getting released for training. The major problem being workload issues, which affected all types of training, and not only the engineers' quality improvement programme.

As a barrier to workplace application of the programme methodologies, UK based Trainees and Trainers perceived that the Trainer-consultants were not being released back into Engineering. Many UK Trainers believed this would make engineers less likely to want to become Trainers for fear of being "left in Education & Training".

Among the Senior Manager and Administrator groups, reference was made to the need for a "critical mass" of trained engineers before organisational change could occur. It was suggested that engineers aged between 30 and 45 who were the "core" of the engineering function constituted the 'critical mass' required to bring about change.

6.2.5.5.5 Organisational Climate
The climate in the company was described by all stakeholders as being uncertain and apprehensive. The majority of interviewees described their enthusiasm for Project 2000 although they were uncertain how it would affect their job roles.

Many stakeholders from the Trainee groups described rumours concerning head-count reduction, and some regarded Project 2000 as potentially "another head-count cutting exercise".

Reader's Note: Project 2000 is a corporate initiative to bring the company's world-wide operations under a global management structure. The aim of the Project 2000 is to produce vehicles for the global markets from common platforms.

6.2.5.5.6 Perception of Education & Training
The perception of the Training department among UK based Managers, Trainees and Trainers stakeholder groups was relatively negative, compared to other activities in the company. Although the introduction of the programme had significantly improved their perception of Education & Training by increasing its credibility (which is reflected in the 1993 feedback), the department was criticised for not fully understanding the needs of its customers. This was due to issues such as inflexibility and course scheduling.

In Germany, the perception of the training department was more positive, with Trainees regarding training as integral to their jobs. However, all Trainees believed the Training department needed to be more customer focused.

6.2.5.5.7 Future Vision
The vast majority of stakeholders perceived the programme to be continuous and that it should be delivered to all engineers in the Company. All interviewees believed that this would be dependent on continued senior management support.

All stakeholders believed there would be extensive use of the technical skills but little change in the use of the behavioural skills until the majority of engineers had attended the programme.

All stakeholders perceived a positive future for the company but recognised that "things would not be easy" over the coming years. Most stakeholders welcomed the challenge of Project 2000 and felt that the programme would be important in increasing the future competitiveness of the Company.

6.2.5.5.8 The Programme in a European Context

The Administrators and Coordinators and Core Team stakeholder groups believed the programme being delivered in the UK differed from that delivered in Germany, particularly with regard to the people skills content. The Germany based Trainers perceived their input into the training design to be insufficient and that some aspects had to be adapted to different cultural circumstances.

6.2.5.5.9 Commitment to the Programme & Workload Issues

Workload was regarded by all to be a constant pressure, with many stakeholders expecting workload pressures to continue to be an issue. The increase in workload pressures was perceived by Trainees and Managers to have some impact on "discretionary" training (such as the engineers' quality improvement programme). Project 2000 was perceived by some as likely to increase their workload, however there was uncertainty as to what specific changes Project 2000 would bring.

The majority of Trainee and Manager interviewees had only participated in the Programme Foundation module. Most perceived it unlikely that they would complete all modules.

Few Managers or Trainees perceived the importance of attending the programme modules in sequence and regarded the modules as individual courses. Many of the interviewees from the Manager Stakeholder group believed that flexible access to the modules would encourage more engineers to attend training which was applicable to their daily work.

6.2.5.5.10 Perceptions of Programme as a Change Programme

Within the Trainees and Managers Stakeholder groups, the programme is perceived primarily as a technical skills training programme. There is little recognition that the programme is concerned with behavioural skills with the manager stakeholder group. The members of the Senior Manager group regarded the people skills as an important aspect of the programme. Among the Trainer, administrator / coordinator groups, together with some members of the Trainee group, the programme' technical and people skills components were perceived as an integrated part of the programme.

All Managers interviewed indicated they would not participate in modules other than the Programme Foundation. This was mainly due to them already have the technical skills either from experience or previous training.

With respect to application of the skills to the engineering process, all interviewees from the Trainee group described problems in applying skills. Some perceived their managers to not understand some of the technical skills and how they should be applied. All Trainees regarded the

Similarly, the following quotations (Table 6-XXVI) are extracted from 1995 interview transcripts with the customers of the programme; (i) the trainees; (ii) managers of trainees; and (iii) senior managers.

Stakeholder Group	Significant Quotation
Trainees:	*(Referring to the flexibility of the programme delivery) "Training occurs in an all or nothing fashion"* *"The relevance of the programme depends on previous training"* *"The programme is UK designed"* *"The programme has to be one of the best and most professionally run courses I've ever been on ....it's an excellent product"* *"The mood in the Company is currently one of uncertainty and apprehension...everybody is unclear as to how their job will change with Project 2000 ... it would be good if we had some clear communication of what is expected, or at least some time frame in which we can expect some answers"* *"Over time you tend to give up trying to apply the people skills bit - it's not part of everyday business"*
Managers of Trainees:	*"We can count the number of engineers going through the programme..but this doesn't relate to the actual application of the skills"* *"The programme needs to be taught in a different style for Germans; its almost a criticism to tell a German that they need team building"* *"Passing the wand isn't what German engineers need"* *"The programme is an under-resourced and inflexible system"* *"Realistically we need 70-80% of our engineers through the programme to see real change"* *"Education & Training need to understand their customer base better to understand our current needs"* *"I've covered most of what is done in the programme in previous training - I do QFD all the time"*
Senior Managers:	*"The programme is too inflexible and if this isn't addressed then we will look elsewhere to train our people"* *"We need to look at ways in which we can ensure the application of the people skills"* *"It is necessary to plan now for a global programme"* *"The programme is going to have to be flexible .... it has to be 'nimble' to respond to change and progress"* *"The programme is important to the long term health of the organisation; I believed that 5 years ago and I believe it now"* *"All the skills should be in common usage - we need to change the mind set"*

Table 6-XXVI: Programme Customer Quotations (1995)

6.2.5.7 Conclusions and Recommendations (1995)

From the 1995 interviews, similar conclusions can be drawn to those from the 1993 stakeholder interviews. They concerned (i) programme outcomes; and (ii) programme content and structure.

(i) Programme Outcomes:

The general perception of the programme by all stakeholder groups (UK and Germany) was positive and all believed that it was beneficial to the future of the company. Managers and Trainees perceived there to be ongoing support for the programme's implementation, and all stakeholders believed the programme contributed to quality improvement and increased competitiveness.

(ii) Programme Content and Structure:

The content of the modules was considered to be appropriate and balanced between technical and people skills, however 1993 comments were echoed with regard to the difficulty in transferring people skills to the workplace. Similar reasons were given in that the predominant organisational behaviour did not readily support teamwork.

Interviewees who had participated in the training enjoyed the training and the style of delivery, however all customer stakeholders and some trainers believed the structure of the programme to be inflexible because of the sequencing requirement for attending the modules. The programme had been designed with each module as a progression from the previous module and as such many of the topics in the later modules required prior knowledge from earlier modules.

The recommendations made to the Programme Steering Committee from the 1995 stakeholder analysis were similar to those of 1993 and are given in Table 6-XXVII.

Action #	Recommended Actions
1.	It is recommended that the programme is delivered to other functions within the Company and to employees at lower levels within the organisational structure.
2.	External suppliers should receive the training so that all levels of the organisation and all functions are able to communicate with a common understanding of both the technical and behavioural skills.
3.	To address the problem of the perceived lack of flexibility in the sequencing of the training programme, there should be more advertising about the programme in this respect

Table 6-XXVII: Recommended Actions (1995)

The recommendations were made to the programme steering committee in November 1995. With regard to recommended action 1, the programme was already being delivered to trainees from areas related to the engineering process. The issue of delivering the training to production supervision was considered at some length, however it was decided that this was not feasible primarily due to issues of release of trainees (Brittle, 1995).

Recommendation 2 had previously been adopted to some extent in recognition of the suppliers' role in the engineering process. It had previously been decided that all quality related supplier training offered by the company, particularly in the UK, would be consistent with the programme. Supplier training was managed by a separate department and was largely out-sourced to external training providers. The recommendation was, however, largely misunderstood or not fully accepted by the Programme Steering Committee. The people skills aspects of teamwork and communication, and the perceived benefits by trainee and trainer stakeholders of engineering process teams undergoing training together, were not provided for by separate supplier training.

Regarding the perceived inflexibility of the training due to the sequential nature of the modules, opinion within the steering committee was divided with members of the development team advocating the sequential nature of the programme and the delivery organisations arguing for greater flexibility. It was decided to take actions to make trainees and trainees' managers aware of the purpose of the sequencing. Information reflecting this need was included in pre-programme administration materials and as part of the programme Foundation module.

6.2.5.8 Attributional Analysis Results from 1995

From the 1995 interview transcripts, 419 attributions were extracted. Of these, 218 were coded as referring to positive outcomes and 201 as referring to negative outcomes.

	Positive Attributions	Negative Attributions
Actual related attributions	141	184
Potential related attributions	77	17

Table 6-XXVIII: All Attributions (1995)

Referring to the programme as an actor , 146 attributions were identified (see Table 6-XXIX).

	Positive Attributions	Negative Attributions
Actual related attributions	57	51
Potential related attributions	33	5

Table 6-XXIX: Programme as Actor (1995)

From the table, the programme is perceived to cause as many positive as negative 'actual' outcomes (57 attributions are positive and 51 negative). Regarding the future, the programme is perceived more positively with potential related attributions referring to more positive than negative outcomes.

With regard to Company as actor, 97 attributions were extracted from the transcripts, of which 30 were positive / neutral and 67 were negative (see Table 6-XXX).

	Positive Attributions	Negative Attributions
Actual related attributions	22	62
Potential related attributions	8	5

Table 6-XXX: Company as Actor (1995)

From the table, there are relatively few attributions which refer to potential outcomes. From the actual related attributions, interviewees perceived a negative effect by the company on the organisation's future.

A cross national comparison between interviewees from the UK and Germany is given in Table 6-XXXI. Of the 419 attributions identified from the transcripts, 259 were taken from transcripts of UK based interviewees and 160 from those of German based interviewees.

	Positive Attributions	Negative Attributions
UK based Interviewees attributions	106	153
Germany based Interviewees attributions	112	48

Table 6-XXXI: Cross National Comparison (1995)

The German and UK interviewees made similar numbers of positive attributions, however the UK based interviewees made more negative than positive comments and significantly more negative comments than their German counterparts. This suggests that overall there is a more negative perception of the organisation in the UK than in Germany.

Analysing cross national comparisons further (Table 6-XXXII), 90 attributions for the programme as actor were positively coded. Of these, 57 related to what interviewees perceived to have actually happened and 33 to potential events.

	UK-based Attributions	Germany-based Attributions
Actual related attributions	26	31
Potential related attributions	20	13

Table 6-XXXII: Cross National Programme as Actor (1995)

From the data, the UK-based interviewees made significantly more positive comments about the future potential of the effects of the programme than their German counterparts. Considering this information with that drawn from the content analysis, the data suggests that UK-based stakeholders are more likely to regard the programme as an ongoing change programme which will yield results in the long term, whereas the Germany-based stakeholders are more likely to regard the programme as a training course which provides immediate skills yielding short term results

6.2.5.9 Conclusions from the Attribution Analysis

The results from 1995 indicated that employees see an optimistic future for the Company and the programme.

Although a large number of attributions were generated from these interviews, the small number of individuals included in the study will mean that care should be taken before extrapolating from the findings. It is clear, however, that they point to interesting and potentially valuable differences between the UK and German engineers at least in this sample. Consequently a larger study investigating the consistency of these cultural differences may well be worthwhile.

All of the interviews were conducted in English and while this could have influenced the way in which German engineers presented themselves, the primary language in the Company is English. As such these differences could still hold important consequences for international communication. Moreover, the finding that these German engineers 'talked' more positively than the UK engineers may also reflect a cultural difference. This could contribute to the German engineers being perceived as more successful. In contrast UK engineers Ford may 'talk' themselves down and in focusing upon negative rather than positive outcomes generate further negative affect.

Attributional'analysis is a relatively time-consuming procedure, but it is one of the only methods available for comparing cross-cultural samples of discourse. By quantifying qualitative material it is possible to identify and differences between groups and track their change over time. A larger group would enable further statistical comparison to be carried out.

6.3 Summary

The programme has been the subject of an extensive evaluation undertaken during its development and implementation stages. The initial conclusions drawn from the pilot evaluation with respect to the programme content and structure and the participants' positive reactions to the programme have been reflected in its implementation.

The implementation evaluation modelled on Kirkpatrick's (1959a) framework has provided some evidence of improvements in participant's knowledge, attitudes, and skills as a result of the programme, and from the stakeholder interviews there is much evidence to suggest that the programme has been successful in terms of its results to the engineering process and the organisation.

Caution is advised however, due to the relatively small sample sizes for the knowledge, behaviour and effectiveness studies and further data collection is warranted for conclusions to be drawn with a high degree of confidence. Furthermore, indications of the programme's effectiveness in improving the engineering process and organisation is based on the perceptions of stakeholders and not on empirical evidence of improvement to the engineering process.

Part III

Meta-Evaluation and New Directions

Part III: Meta-Evaluation and New Directions

> *'New opinions are always suspected, and usually opposed, without any other reason but because they are not already common.'*
>
> *- John Locke, 1690*

From the empirical study and with reference to the reviews of the literature in Part I, conclusions are drawn with respect to the utility of training evaluation. A critical review of the empirical study is given in Chapter 7. The basis for this review is from the perspective of a participant-observer, with an interest in the development and improvement of training and the meta-evaluation is undertaken against a four dimensional criteria of; utility; feasibility; propriety; and accuracy. General conclusions are reached from the meta-evaluation within the context of inferences drawn from the reviews of the literature in Part I of this study.

Chapter 8 consolidates the conclusions from chapter 7 and considers a parallel conception of evaluation in the industrial and commercial contexts. The evolutionary history of evaluation in this setting provides directions for the development of the evaluation of training which addresses the problems identified through parts I and II of this study.

Chapter 7

Review of the Empirical Study

7. Review of the Empirical Study

> *'Learning from experience is tough; you get the test first and the lesson afterwards.'*
>
> *- Dr Peter Honey (1997)*

7.1 Introduction

This chapter provides a critical analysis of the evaluation approach employed for the programme. The analysis is directed by the research focusing question 1 (chapter 2: section 2.6.1) 'what role does training evaluation, conceived in terms of Kirkpatrick, play in training improvement within a commercial context?' and is based on the specific experiences of the empirical study. Consistent with the broader intent of this research, the analysis is not restricted, however, to just the Kirkpatrick elements of the empirical study.

The evaluation methodology, together with the use of the resultant data in the management of the programme, are reviewed and analysed and the overall evaluation philosophy is examined. The conclusions drawn from this analysis provide the central theme for this thesis; an examination of the underlying philosophical foundations of training evaluation, which leads into chapter 8 which is focused by the second research question (chapter 2: section 2.6.1) 'what restructuring is necessary for training evaluation to integrate with the theories and practice of learning and training?'

7.2 Meta-Evaluation Methodology

The application of evaluation to itself is sometimes called meta-evaluation (Scriven, 1996). Stake (1973) argues that meta-evaluation provides quality control for evaluation activities. Basarab and Root (1992) argue that meta-evaluation with respect to training, is concerned with assessing whether the evaluation has provided the 'best possible information' defined in terms of quality, usefulness, feasibility and technical accuracy.

7.2.1 Meta-Evaluation Design

The design of a meta-evaluation is determined primarily by its purpose (i.e. what is it that needs to be known; why is a meta-evaluation deemed to be necessary). As introduced in the previous paragraph, purposes of meta-evaluation recognised in the literature include quality, usefulness, feasibility and technical accuracy dimensions. With reference to this study, the purpose was defined: 'this programme provided an opportunity to apply and study training evaluation within its contemporary paradigm with the purpose of providing a directional framework for training evaluation within commercial organisations' (Introduction: paragraph 1). This purpose statement prompts a broader meta-evaluation design which considers the underlying assumptions of training evaluation, as well as the specifics of the evaluation itself.

The study purpose statement is comprised of five components; (i) apply; (ii) study; (iii) contemporary paradigm; (vi) directional framework; and (v) commercial organisations. Each component can be elaborated upon to facilitate the design process.

The application component is concerned with the design of the evaluation and its constituent instruments (i.e. Kirkpatrick's framework and the various data collection and analysis techniques). Application also includes how the evaluation was deployed to the engineering training programme (i.e. feasibility and propriety of evaluation instruments), who was involved in collecting, analysing and communicating evaluation information (i.e. ownership and reporting) and how the information was used (i.e. utility).

The study component is concerned with; a) the informed observation of the application of the training evaluation; and b) analysis of observations against criteria for the given meta-evaluation purpose. With respect to a), informed observation is contextual, with the context being learning, training and the corporate setting. Learning and training have therefore been subject to an extensive, although by no means exhaustive, review in chapter 1, with informative supplemental text given in the appendices. The corporate training setting has also been the subject of a detailed, but again not exhaustive, account in the introduction and in chapter 4. Supplemental text which is illustrative of the setting and the engineering process and process improvement methodology is also included in the appendices.

The contemporary paradigm component recognises that members of a scientific community share a paradigm (Kuhn, 1970) and that paradigms shape and constrain the conceptualisations of that community (Patton, 1990). The reviews of the literature of the conceptualisation of evaluation, particularly with reference to training evaluation (chapter 2), have illustrated the nature of the contemporary paradigm. There is, of course, some divergent persuasions within the literature which have evolved from experience and thinking, but they have been largely held to the general boundaries of the paradigm.

The significance of paradigms cannot be understated with respect to this study. There is sufficient evidence in the literature to suggest that the evaluation of training is failing the commercial sectors of society. As previously stated, the literature on training evaluation is voluminous and yet much of it is isolated from practitioner activity. Furthermore, development of evaluation thinking is internal to the current conceptualisations (particularly with respect to Kirkpatrick), with evaluators paying attention to validity and reliability issues surrounding data collection and analysis. Recognising and challenging the contemporary paradigms of training evaluation is therefore an important aspect of this research.

The directional framework component relates to the paradigm component of this research in that it is intended to outline some guiding principles or sign-posting for future research. This aspect of the study is not intended to prescribe a model for training evaluation in any significant degree of detail. A directional framework offers research and practitioner agenda opportunities for training evaluation. In this sense, the framework facilitates the development of research questions to move our thinking on training evaluation into a new era.

Finally, the commercial organisation component of the purpose statement brings into focus the nature of training in organisations where business success is the ultimate driving factor. Training in this context serves to improve business performance and is largely dependent on its contribution to the business for its long term survival.

Given the wide ranging nature of the study purpose statement, the design requirements of the meta-evaluation are two dimensional. The first dimension reflects the evaluation in terms of its general design characteristics (i.e. the evaluation conceptualisation, purpose and communication processes, together with its multiple data collection and analysis techniques). The second dimension is the criteria against which each of these components are contrasted. The criteria dimension is developed in section 7.2.3.

7.2.2 Informed Observation

A further design consideration of this meta-evaluation is its context to the relevant bodies of knowledge in the literature. The meta-evaluation therefore draws on the reviews of the literature in Part I; learning and training (chapter 1); conceptualisations of evaluation (chapter 2); and evaluation methodology (chapter 3), which are set in the context of the training programme (chapter 4).

The primary source of empirical data is provided by the researcher's own research diary (research log) which comprises of entries made over a five year period of meeting notes, observations, and discussion notes, together with the researcher's own documented thoughts which originated during the course of designing, developing, and implementing the training programme. The entries were made real-time by the researcher as he worked with programme staff, management, and trainees.

Dane (1990) comments that the research log or field journal is the research tool of greatest importance to the participant-observer. He notes that entries fall into two categories of certain and uncertain and that it is essential that these are distinguished by the researcher. Forgas (1982) illustrates this by providing evidence of 'blank filling' by recorders of observations where actions are logically inferred from observations, but not directly observed (e.g. if a subject puts his coat and hat on and leaves the classroom, an observer may infer that he is going out of the building even though he was only observed to have left the room). Dane also comments that field journals should be maintained as often as possible.

With respect to the compilation of the research log of the empirical study, records of observations were distinguished from thoughts and perceptions and in this sense were recorded as certain and uncertain. Direct observations were recorded as such and care was taken to ensure that any inferences drawn were identified as such. The log was maintained frequently, often during key stages or events within the programme.

7.2.3 Meta-Evaluation Criteria

From the study purpose statement elaboration in section 7.2.1, four interest themes emerge; a) the usefulness of the evaluation information provided; b) the practicality of using the evaluation design and its constituent instruments in the training programme and the industrial setting; c) the legal and ethical implications of evaluating training; and d) the extent to which evaluation information reflects reality. These interest themes form the basis for the meta-evaluation criteria requirements.

These requirements fall within the criteria guidelines proposed by Scriven (1972) and Stufflebeam (1975). Scriven and Stufflebeam provide criteria for meta-evaluation and these are reflected in the 1981 Standards for Evaluation of Educational Programs, Projects, and Materials published by the US Joint Committee on Standards for Educational Evaluation. Stufflebeam and Madaus (1996) observe that the most important feature of the Joint Committee was the representative nature of its membership which ranged from technically oriented groups to a variety of practitioner oriented groups. Basarab and Root (op.cit.) endorse the use of the standards as a basis for criteria for meta-evaluation of training, observing that they are applicable to a range of evaluations varying from small to large-scale, formal to informal, and formative to summative.

The published set of standards consist of thirty individual standards which are grouped into four categories of; (i) utility; (ii) feasibility; (iii) propriety; and (iv) accuracy. For the purposes of this meta-evaluation, it is not necessary to examine each of the thirty individual standards, but to review the general categories as the basis for developing an organising framework for the meta-evaluation.

7.2.3.1 Utility Criteria

Utility standards are those which are intended to ensure that an evaluation will serve the practical information needs of given audiences. The standards within this group are concerned with the identification of audiences involved in or affected by the evaluation, the scope and type of information, and the nature of communication of evaluation information. With regard to the audiences of evaluation, Basarab and Root (op.cit.) identify those audiences within a corporate training context to typically include training managers, instructional designers, course developers, participants, participants managers, funding organisations and corporate councils or training boards.

The scope and type of information is defined in terms of the pertinent questions as determined by the programme stakeholders about the evaluand and is therefore responsive to the needs and interests of specified audiences. The interpretation of value conveyed by the information is subject to a range of perspectives and procedures and the value criteria is therefore likely to vary between audience groups.

With respect to the nature of the communication of evaluation information, the standards are concerned with the clarity of information reports, to whom the reports should be made and the timeliness of the reporting. The joint committee require communication (reporting) to be timely so that audiences can make best use of the information.

The standards conceive communication in terms of reports; formal written accounts which convey information. This reflects the evaluator - client conceptualisation of evaluation which is consistent throughout the literature. For the purposes of this study however, guidelines provided by the standards have been generalised to the activity of communication. This does not presuppose the evaluator - client relationship. The significance of this is discussed in section 7.4.1: Ownership of the Evaluation.

7.2.3.2 Feasibility Criteria

Feasibility standards are those which are intended to ensure that an evaluation will be realistic and diplomatic..The standards in this group are concerned with practical procedures, political viability, and cost effectiveness. Practical procedures refer to evaluation as procedure and require that disruption from evaluation activities is kept to a minimum and that the information needed can be obtained.

Basarab and Root (1994) comment that training evaluation activities should be planned to complement and enhance learning experiences as opposed to hindering them. Easterby-Smith (1994) observes the significance of the role of the Hawthorne effect (Mayo, 1946) in this respect, noting that it may often be difficult to distinguish between actions undertaken as a result of the evaluand as opposed to as a result of observation or interest as part of an evaluation.

Political viability recognises the different positions of various interest groups (department and organisations within the company) so that their co-operation may be obtained, and to avoid attempts by these groups to obstruct evaluation activities or to bias or misapply the results. The cost effectiveness standards advocate that evaluations should produce information of sufficient value to the audience so as to justify the resources expended. Again, this aspect of the standards reflects the evaluator - client paradigm, with the evaluator offering value for money of his/her services to the client. However, the general principles apply to any form of evaluation, in that the cost of obtaining data should not outweigh the benefits of having the data.

7.2.3.3 Propriety Criteria

Propriety standards are concerned with the legal and ethical conduct of evaluation activities. The scope extends to those involved in the evaluation and those affected by its results. The standards reflect the review of the literature with respect to evaluation ethics (chapter 3). The standards which constitute this group consider obligation, disclosure and human rights. Obligation refers to the roles and responsibilities for evaluation activities. The standards associated with this set out the need for those involved to fully understand their responsibilities and to commit to fulfil their obligations. Any conflicts of interest should be dealt with openly and honestly so as not to compromise the evaluation processes and results.

Disclosure is concerned with standards for written and spoken evaluation reports, advocating they be open, direct and honest in their disclosure of findings, including the limitations of the evaluation. Standards for human rights, which relate to disclosure, are concerned with respect for human dignity and worth requiring evaluation activities to be designed and conducted so that the rights and welfare of the human subjects are respected and protected. The standards also concern the need for balanced communication (reporting) which is fair in its presentation of strengths and weaknesses of the evaluand.

7.2.3.4 Accuracy Criteria

Accuracy standards are intended to ensure that an evaluation will provide sound information. Stufflebeam and Madaus (1996) describe sound information as information which is technically adequate and features conclusions that are linked logically to the data. Accuracy standards included in this group are concerned with; object (evaluand) definition and context; measurement; and conclusions.

The definition and context standards refer to the clear identification of the attributes of an evaluand which are of interest and calls for the examination of the content, purpose and business requirements. The wider context of the evaluand is examined for probable influences on the evaluand so that they are figured into the evaluation strategy.

Measurement standards are concerned with validity and reliability of data collection and analysis. Conclusions drawn from evaluation data are subject to standards of explicit justification and defensibility, with safeguards within the reporting process against personal biases and distortions by any party to the evaluation.

7.2.4 Organising Framework

The standards are widely acknowledged (Basarab and Root, 1992; Madaus et al, 1996; and Shadish et al, 1995) to provide a comprehensive criteria dimension of best practice. Basarab and Root (1992) have refined the standards into a meta-evaluation instrument intended to determine whether an evaluation worked in terms of the organisation's needs and to identify areas of improvement. For the purposes of this review, the four general categories form the basis for the meta-evaluation framework.

The empirical study comprises a multitude of evaluation instruments, with each applied in a distinct manner by means of the way data is collected and analysed. Each instrument is designed to meet a specific evaluation need. The instruments are selected to meet the overall purpose of the evaluation. Each instrument has its own distinct communication process which contributes to the overall communication component of the evaluation. With respect to the implementation evaluation, the conceptual framework by which the evaluation is conceived is a significant component for review.

The evaluation therefore can be described in terms of specific and general components. Each evaluation instrument is a specific component of the evaluation in that it is conceived to meet specific needs. Evaluation purpose, communication, and the implementation evaluation framework are general components of the overall evaluation in that they relate to each of the specific instruments.

7.2.5 Review methodology

For the purposes of this review, each specific component of the evaluation is analysed in terms of all four criteria from which conclusions about the feasibility, propriety and accuracy of the overall evaluation can be inferred. The organising framework of value criteria is intended to serve as a guide and provide structure to the review. For conclusions with respect to utility, it is necessary to review the evaluation purpose, communication processes and overall framework (Table 7-I).

Evaluation Components		Meta-Evaluation Criteria			
Pilot Evaluation	Implementation Evaluation	Utility	Feasibility	Propriety	Accuracy
Training Observation		✓	✓	✓	✓
Reactions (spoken feedback)		✓	✓	✓	✓
Focus Grp. Feedback Questionnaire		✓	✓	✓	✓
	Level 1: Reactions (spoken feedback)	✓	✓	✓	✓
	Level 2: Knowledge Gain Questionnaires	✓	✓	✓	✓
	Level 3: Attitudinal Questionnaire	✓	✓	✓	✓
	Level 3: Behaviour Observations	✓	✓	✓	✓
	Level 4: Stakeholder Analysis (Interview)	✓	✓	✓	✓
Purpose of Evaluation		✓			
	Purpose of Evaluation	✓			
Communication		✓			
	Communication	✓			
	Kirkpatrick Framework	✓			

Table 7-I: Meta-Evaluation Organising Framework

As a member of the programme development team and of the evaluation group studying the programme, my position with respect to this analysis is one of a participatory evaluator. Patton (1990) describes participatory evaluation as a reflective process by the person(s) in the programme (evaluand) or community for their own development. This meta-evaluation is heuristic in its nature as personal experience and insights of the researcher yield an understanding of the phenomenon of interest. Douglas and Moustakas (1984) describe heuristic inquiry as being concerned with meanings, not measurements; with essence, not appearance; with quality, not quantity; and with experience, not behaviour.

The analysis of the programme draws from observations and thoughts recorded in a research log. This log was used to document the researcher's own observations, and those of others, throughout the implementation of the programme (1991 - 1995) together with the researcher's thoughts at that time as to how the programme and its evaluation were progressing and how it might be improved.

7.3 Analysis of Evaluation Components

A range of data collection methods (Table 7-II) were used as part of the pilot evaluation and of the implementation evaluation.

Evaluation Phase	Data Collection Methods
Pilot Evaluation	Training Observation Reactions (spoken feedback) Focus Group Feedback Questionnaire
Implementation Evaluation	Level 1: Reactions (spoken feedback) Level 2: Knowledge Gain Questionnaires Level 3: Attitudinal Questionnaire Level 3: Behaviour Observations Level 4: Stakeholder Analysis (Interview)

Table 7-II: Evaluation Instruments

7.3.1 Pilot Evaluation

The pilot evaluation was comprised of four specific components; observation and reactions; focus group; and feedback questionnaire and two general components of; purpose; and communication.

7.3.1.1 Observation and Reactions

Although the training observation and trainee spoken feedback data collection methods employed for the pilot evaluation were largely informal and undocumented, both were observed to have a direct and immediate impact on the design of the programme. With respect to utility, the insights into trainee's perceptions of the training at various stages of each module provided the training development team with a clear indication of where improvement actions were necessary to the instructional design. The feedback to the programme developers was real-time and direct and the spoken feedback complemented the observations making explicit the thoughts and opinions of trainees.

As a pilot programme, members of the development team concerned with the particular module were present throughout and in this sense observing participants and receiving spoken feedback were feasible, although it should be noted that this level of observation of trainees would be infeasible (i.e. cost of observers) for regular ongoing training. The ongoing feasibility of spoken feedback is discussed as part of the meta-evaluation of level 1 implementation evaluation (see section 7.3.2.1).

With respect to propriety, the members of the development team were essentially participants in the pilot; the role of the pilot programme was declared to trainee engineers as being for development purposes. Trainees were fully aware of the use of observation and why this was necessary. End of day and end of module spoken feedback was not recorded. This was intentional as it was believed that it would inhibit some trainees when making comments. After each day of a module, the development team reviewed their interpretations of observations and spoken feedback and agreed action plans for the improvement of the course with respect to content, timing, presentation of materials, and the dynamics of learning activities.

7.3.1.2 Focus Group

The end of pilot programme focus group was semi-structured allowing participants to identify positive and negative aspects from an undefined range. The levels of co-operation which had been developed between the trainer (development team) and trainee group gave a high level of participation and 16 of the final 22 pilot trainees attended.

The data provided with respect to areas of strength and improvement was free flowing and highlighted key areas in terms of; marketing sales points for the programme; and aspects which were important to trainees and required improvement, and was therefore useful to the programme management.

With respect to propriety and accuracy, similar observations can be drawn to those for spoken feedback and observation.

7.3.1.3 Feedback Questionnaire

The questionnaire was designed to explore areas of the programme which were important to the development team in that they were unique to the programme or were considered to be important with respect to the programme's potential to bring about change in the engineering process. In this sense, the questionnaire fulfilled its utility requirements by giving specific, timely and direct feedback on those characteristics.

The questionnaire was administered as part of the focus group session and as such was a feasible measurement instrument to apply. Regarding propriety and accuracy, the questionnaires were administered anonymously and sought trainee perceptions of the training. Analysed in the context of the focus group feedback and with the opportunity to explore any aspect further, this approach can be considered to be valid and reliable for its purpose.

7.3.1.4 Purpose of the evaluation

The purpose of the pilot evaluation was primarily to; assess the viability of the instructional design; gauge engineers' reactions to the technical and people skills content; and assess the relevance of the programme to the engineering process (chapter 5: section 5.2.1).

In terms of its utility, the pilot evaluation provided practical information to the evaluation audience; the development team and the programme steering committee. By the nature of its observational element, the development team were able to receive direct detailed feedback, which was supplemented by the spoken reactions, allowing them to capture the nuances of the programme design with respect to its instructional approach and content.

The pilot evaluation was limited however in that it did not explore the outcomes of the programme in terms of its impact on the engineering process. This could only be hypothesised from the reactions and comments made by the pilot study engineers. Another observation of significance is concerned with the 20%/80% programme development process; where in terms of cost and time 20% of the overall development occurred prior to the pilot and the remaining 80% came afterwards. This extensive post-pilot development of the programme was not directly as a result of the pilot study data. The continuing redevelopment of the programme was observed to be undertaken largely as a result of opinions of members of the development and delivery teams without consideration of any evaluation data.

7.3.1.5 Communication

The communication process was direct and timely in that the pilot programme trainees were observed by, and gave feedback to, the development team. In the context of the pilot, the development team were the primary audience for the evaluation data. Practical and timely actions were facilitated by this communication process.

The sole use of spoken feedback as a measure of trainee reactions, whilst being timely and efficient, does not facilitate communication to wider audiences. The absence of written data prevents the dissemination of accurate reactions information; relying on the interpretation of the development team in conveying the message.

7.3.2 Implementation Evaluation

The implementation evaluation comprised of five specific components; reactions; knowledge gain questionnaires; attitudinal questionnaires; behavioural observations; and stakeholder analysis interviews and three general components of; purpose; communication; and Kirkpatrick's framework.

7.3.2.1 Level 1: Reactions to the Programme

The use of spoken feedback as part of the interim and final stages of each module provided all participants with the opportunity to express their reactions to the training. Regarding the utility of this type of feedback, the information is given directly to the trainers and as such, non-verbal reactions to the training can be observed. The trainer has the opportunity to immediately follow-up reactionary comments and explore the causes of a particular reaction or group of reactions.

The feedback is real-time in that the information is collected and assimilated immediately allowing trainers to act on feedback where considered appropriate and feasible. The trainer is given the opportunity to listen to and understand concerns and is empowered to take the necessary action. The trainees are given the opportunity to express their reactions to the training with the knowledge that their feedback is being listened to. This type of feedback does not generate masses of documentation requiring analysis and reporting (often back to the trainers who collected the data).

For the trainees, being required to express their reactions to a programme in front of other trainees and directly to the trainer may result in trainees being less critical. However, from stakeholder feedback (see section 6.2.5 Programme Effectiveness) which was provided during one to one interviews with an outside agent not associated with either the development or delivery of the programme, trainee comments reflected much of the spoken reactions feedback obtained during the training.

The absence of documented evidence did not allow those who were not present at the feedback sessions to receive un-interpreted data.

This type of data collection method has the potential for the comments of one trainee to influence the subsequent comments made by other trainees. However this was not perceived to be the case by the trainers who received the feedback.

7.3.2.2 Level 2: Changes in Knowledge

The pre- and post- module training questionnaires were administered and analysed by the trainers, with feedback on knowledge gain being given to participants prior to them leaving the training.

The questionnaires were initially conceived for the Programme Foundation module by the development team as a means to demonstrate knowledge gain to members of the two executive technical groups as part of the programme approval process. The design template was subsequently applied to the development of the pre- and post- knowledge questionnaires for all modules.

As part of the development of the measure, the Programme Foundation module questionnaire was revised to include parallel forms of the question items to eliminate gains due to memory. Whilst these changes were proven to be successful in terms of the reliability of the measurement instrument, the improvements were not transferred to the knowledge gain questionnaires applied to the other modules.

The scope of the knowledge assessment, an aspect of utility, was limited. Emphasis had been placed on the development of the knowledge assessment for the programme foundation module and a valid and reliable test had been achieved. However this design was not transferred to the other modules and as such, no reliable data could be obtained for those modules with respect to knowledge gain.

This was observed to be largely due to the positioning of the evaluation responsibility in the organisation and the subsequent lack of involvement of the majority of the delivery teams. No data was obtained from Germany, France, Spain or Portugal. In the case of Germany where frequent contact was maintained, trainers were reluctant to co-operate with the evaluation effort, although it has been reported via the programme steering committee that the pre/post knowledge questionnaires were regularly used in the German modules. In the case of France, Spain and Portugal, the remoteness of these operations and the general lack of communication is considered to largely contribute to their lack of involvement in the evaluation effort.

The non-integration of the evaluation into the delivery of the programme, particularly with respect to ownership, is a recurring theme throughout this study and as such the whole issue of ownership is considered later in this chapter.

The post- course knowledge questionnaire and the feedback session which were integral to the closing stages of each module provided participants with feedback on their learning progress. The role of the knowledge assessment was therefore twofold; (i) to provide the programme staff with feedback on the learning gain; and (ii) to provide feedback to trainees.

The participant identification system allowed participants to undertake the knowledge questionnaire without being identified and therefore with minimal personal risk. Scriven (1996) describes this phenomena as 'valuephobia' - a pervasive fear of being evaluated. As quality methods form part of an engineer's job competencies (abilities required to perform the job competently), participation in any form of competence assessment provides a measure of an individuals ability. By using the participant identification system, individuals were ensured anonymity. The pre- post- knowledge evaluation had a 100% return rate from those who completed the training module.

With respect to propriety, the participant identification whilst providing a means for making individual pre- post- module knowledge did not fulfil it's potential to track participants' knowledge gain throughout the entire programme. This was observed to be largely due to two factors; (i) a high number (>30%) of participants failed to recall their personal identification numbers at subsequent modules; and (ii) a central data-base of knowledge gain was not maintained. With respect to (ii), once again this was observed largely to be due to ownership of the evaluation effort by the delivery team.

The knowledge questionnaire was an integral part of the course design. It was observed that the pre- knowledge instrument had the effect of sensitising participants to key concepts in the programme (Brittle, 1991-1995). It also provided participants with an indication of their own level of knowledge with respect to the company's quality philosophy and methodologies to support the engineering process. This was particularly relevant to the people skills aspects of the programme.

From the description of the content and structure of the training in chapter 4, each module was designed in terms of behavioural learning objectives according to Bloom and necessarily providing training in a range of skills and to a variety of levels. The feasibility of administering the pre- and post- evaluation questionnaire determined each to last no more than 30 minutes. This represented approximately 5% of the training time.

To accommodate these time constraints, 25 multiple choice question items were used. From figure 5-VI: Content / Question comparison matrix in chapter 5 which describes only the general objectives, it is not feasible in the time available to subject all objectives to a question item. A sample of objectives are therefore selected for question items from which the knowledge gain for all objectives is inferred. As a single set of parallel form pre and post questionnaires was used, a tendency was observed amongst some trainers to place particular emphasis on those aspects of the training which were the subject of a question item. This emphasis on sample item related objectives by trainers is hypothesised to cause trainees to gain a better understanding of this material and therefore perform better in the knowledge tests. As the sample is used to infer knowledge gain for the entire module but is not representative of all objectives, the inference cannot be considered to be valid.

7.3.2.3 Level 3: Changes in Attitudes

Attitudes were evaluated in terms of trainees' perceptions of the organisation and of quality using a combined survey questionnaire comprising the two instruments. For the purposes of this review, these two evaluation instruments are considered separately and general conclusions for the components are reached.

7.3.2.3.1 Perceptions of the Organisation

The utility of the questionnaire in this respect is concerned mainly with the practicality, scope and type of information gained. Perceptions of a supportive organisation, whilst generally considered to be an important aspect to the programme, are not a function of the programme and therefore beyond the scope of the programme aims. The information gained was contextual in that it provided a measure of some of the environmental factors within which the programme was implemented. Consequently the data provided via the component's communication process (see below) held no value with respect to how the programme could be improved.

This issue of value of the information warrants further discussion; the information provided was observed to be widely regarded as interesting and informative. Indications of engineers' perceptions of the conduct and efficiency of management in performing its role, risks associated with seeking help, and trust in colleagues, provided general insights into the engineering environment, which would perhaps have been useful during the formative stages of the programme as part of a needs analysis exercise. However, as a measure of the programme's effectiveness in changing attitudes, the factors analysed did not relate to the objectives of the programme and could not be used to improve the programme. Basarab and Root (1992) comment that the value of evaluation information, as aspect of accuracy within the meta-evaluation criteria, is determined by the needs of its beneficiaries at the time it is received and therefore this information was of no value to the programme staff or management.

The communication of the data was undertaken using a formative evaluation report made to the programme steering committee. This process occurred during the closing stages of the programme and was therefore too late to enable the significant changes which it suggested to be made. Once again, the issue of timeliness of communication is highlighted as a critical feature in the evaluation effort.

Observations concerning the feasibility criteria for this component are concerned with its administration and the cost of obtaining the data. These two criteria are related; the administration of the component was not integrated into the delivery of the training and was therefore incremental and cost adding. Apart from the costs of the engineers' time to complete and return the questionnaire and despite the use of e-mail, an administrator was engaged in the distribution and follow-up for the return of questionnaires. This work was observed to be additional to the other programme administration duties and subsequently for stages 2 and 3 of the survey, significantly smaller sample sizes were achieved (<15% of stage 1). The cost of the organisation questionnaire can be concluded to be high, particularly with respect to the value of the information provided.

Concerning propriety of the component, although e-mail was used which prevented anonymity, questionnaire data was not attributed to individuals. Trainees' rights in terms of the Joint Committees standards were observed not to be compromised.

Regarding accuracy, four critical sets of observations and conclusions can be drawn from the empirical study. The measures of the perception of the organisation were developed by members of the trainer-consultant team. The exemplars identified (from which the measures were developed) were those of a hypothetical organisation supportive to quality improvement as defined by Deming (1982). In this respect, the measure was concluded to be valid in that it was developed from expert knowledge. However, the target audience for the programme was the operational engineers within the product development and manufacturing divisions. Management of these divisions were identified as only requiring the programme foundation module as their role was not as engineering practitioners, but as managers of the engineering process. From training records for the UK and Germany for the period leading up to stage 3 of the evaluation, more than 70% of engineers had attended 3 or more programme modules whilst less than 20% of all engineering management had attended the programme foundation module.

Likert scales were chosen for the measure mainly because of the knowledge within the evaluation team, however, Oppenheim (1992) observes that to study attitude change, Guttman's method is preferable.

7.3.2.3.2 Perceptions of Quality

Observations concerning feasibility and propriety are as those made for the organisational component, however the cost consideration differs in the context of the value of the information provided by the quality component to the programme staff. The information provided was observed to be of value to the programme staff and management and therefore providing function for cost.

The utility of the data provided is more directly related to the programme than that for the organisational component. The innovation and excitement dimension provides a measure of the underlying aim of the programme; to shift engineers' thinking from a problem solving or corrective engineering paradigm to one of planning and integration of quality in customer terms into the product at the concept or design stage. Similarly customer empathy, creative engineering, product innovation and open to new approaches relate directly to the purpose of the programme.

The communication process was the same as that for the organisational component and therefore the same observations apply. However given the potential utility of the data, timeliness is highlighted as a significant weakness in the evaluation approach, with information provided to stakeholders too late in the programme to take any corrective action.

Concerning accuracy, the final battery of items for each dimension were selected on the basis that their correlation coefficient (Pearson's Product Moment) was greater than 0.7. Content validity was assumed on the basis that they were developed by subject matter experts. With respect to their technical adequacy, the results of the survey (chapter 6; section 6.2.3.3) on the innovation and excitement dimension indicate an initial improvement between stages 1 and 2 with no further improvement at stage 3. Similar results and conclusion were drawn for creative engineering, product innovation and open to new approaches (chapter 6; section 6.2.3.4). This was observed to align with the development team's expectations. However, the results for customer empathy (chapter 6; section 6.2.3.4) suggest no evidence of change. Given that an understanding of the customer is an underlying principle of the programme, this warrants further investigation.

A review of the customer empathy survey items by members of the customer focussed engineering module team concluded that 4 of the items would not allow trainee's to distinguish basic, performance, and excitement quality features (chapter 4; section 4.2.1.7) and 2 of the items did not relate to customer empathy. On the basis of this review, the customer empathy dimension items were concluded not to be technically adequate and therefore not valid.

A further observation was recorded (Brittle, 1996) concerning the relationship between all survey items and programme objectives. The joint committee advocates clear identification of attributes of the evaluand with respect to evaluation data. This was observed not to be so, with the evaluation results being couched in general terms and not indexed to specific module objectives. Whilst further investigation would be likely to reveal these relationships, this would require further surveying with more specific items. In this sense the evaluation results cannot be considered to be technically adequate for the purposes of programme improvement.

7.3.2.4 Level 3: Changes in Behaviour

Changes in behaviour were evaluated using observation rating scales. This element of the evaluation was developmental and was not transferred to the wider implementation of the programme, however observations concerning utility, feasibility, propriety and accuracy were documented.

The potential utility of any measure of behavioural change is high in that it seeks to indicate changes in trainee behaviour resulting from the programme. The underlying purpose of training in a work organisation to bring about changes in behaviour and business related performance (Jackson, 1989) and therefore training evaluation data which indicates a programme's effectiveness in these terms is useful to programme sponsors, managers, and trainers.

The underlying purpose of the engineers' quality improvement training programme is to improve product quality and to this end the programme includes engineering and manufacturing methodologies intended to facilitate improvement. Recognition of the needs for engineers to work together in applying these skills led to the inclusion of the people skills curriculum in the programme. The behavioural rating scales developed to measure behaviour change were designed solely for the purpose of measuring people skills; a secondary or enabling aim of the programme. Although the data provided related to specific objectives within modules of the programme and was therefore potentially useful to programme staff, it gave no indication of behavioural change with respect to the primary technical (quality improvement) skills.

During the design stages of the programme, application check-sheets were conceived as a means to aiding engineers in the application of the technical skills (chapter 5; section 5.3.7.1). These were not embraced by the evaluation group and consequently did not form part of the overall evaluation strategy. Trainers were observed to employ the check-sheet as part of their applications support role (chapter 4; section 4.5.3) by using the check-sheets to facilitate applications and further trainee learning by way of a review process with engineering teams. These reviews were not recorded centrally and therefore a body of knowledge with respect to technical behaviour change was not maintained.

The generality of the people skills increased the potential utility of this type of data, although the limited application did not provide data from a sufficient sample to draw conclusions with any degree of certainty. The nature of communication was a formative evaluation report to the programme steering committee, however those trainers who were involved in the FMEA application for which the behaviour observation rating scales were used commented that the feedback was useful for identifying learning opportunities for the group.

Regarding feasibility of the observation rating scales, the scales were used in two ways; by a member of the evaluation team; and by trainees themselves. Used by the evaluation team member, the observation rating scale was observed to cause minimal disruption to the groups, however trainees reported they were initially conscious of the observer's presence at the meeting but soon became used to them being there. From the data provided from this limited sample, there is no evidence to suggest influence on the data from the Hawthorn Effect (Mayo, 1946). Trainees reported they found self-rating helpful in analysing their own performance, with several commenting that descriptions of negative behaviour provided learning insights for observing and intervening in team process.

Concerning the cost aspect of feasibility, using a member of the evaluation team to collect data is considered to be high and prohibitive on a large scale. At a daily cost of £500 per day (in 1995) and with in excess of one thousand teams participating in the programme, observation of a representative sample of team applications (i.e. engineering, administrative, cross functional, cross national, and multi-organisational level) is not feasible. However, the use of self-rating, if proven to be valid from analysis of a wider sample of applications, has minimal incremental costs and was observed to further learning in team behaviour.

The limited evidence from the empirical study suggests that self-rating is robust to observational errors when conducted using reliable rating scales, however this approach has been criticised in the literature i.e. Golembiewski et al (1976) who distinguish three types of change (in behaviour) resulting from training in terms of self-rating; alpha; beta; and gamma change. Where alpha change is a true change, and beta and gamma changes are due to trainee reconceptualisations of the measurement instrument and constructs of interest respectively. These trainee induced errors, if unquantified, affect the reliability of the measure.

With respect to propriety, the ethical conduct of the observer is of central interest. As a trained and experienced observer was employed for this aspect of the evaluation, ethical guidelines (chapter 3; section 3.4) were satisfied. The role of the observer was explained to trainees (disclosure) and reporting was balanced in that it reflected positive and negative aspects of the application of the programme people skills.

Analysis of the accuracy of the observation rating scales is hindered by the small size of the sample application and therefore the following observations and comments should be regarded as tentative, requiring further empirical investigation to increase confidence in this analysis. In terms of its technical adequacy, it is reasonable to assume the measurement instrument is adequate for its intended purpose in that it was developed by and checked by experienced trainer-consultants

(Chapter 5; section 5.3.7.4). Similarly the attributes of interest were clearly identified. As previously considered the validity of the measure, in terms of the business requirement, was limited as it did not contribute to the measurement of changes of behaviour in terms of the primary objectives of the programme.

7.3.2.5 Level 4: Programme Effectiveness

Stakeholder analysis was employed to elicit the perceptions of key programme stakeholders based on their experiences of the programme. The stakeholder analysis provided qualitative feedback, and although (as previously discussed in chapter 1) this type of data has limited currency in the technically and commercially oriented world of the automotive industry, Rousseau (1990) summarises the worth of this type of qualitative analysis by suggesting;

> *"quantitative assessment offers opportunities for inter-organisational comparisons to assess relations between culture and organisational success, strategy and goals.....qualitative research can explore the meanings behind the patterns."*

In the sense of Kirkpatrick, level 4 evaluation is concerned with results and Kirkpatrick (1960b) classifies these as; reduction in costs; reduction in turnover and absenteeism; reduction of grievances increase in quality and quantity of production; or improved morale which may lead to some form of improvement. The use of stakeholder analysis provides qualitative information with respect to many of these aspects, however it does not provide quantitative information.

The absence of so-called bottom line data (Brinkerhoff, 1991) in this study is a major deficiency. Attempts to evaluate this programme in terms of its cost/benefit ratio (Kearsley, 1982; Phillips, 1991) were considered as part of the level 4 evaluation, however it was concluded that this type of analysis was of high risk to the organisation; costs are relatively easy to estimate, however benefits in financial terms are more difficult and a feasible and valid method of collecting this type of data was not identified. By reporting the costs of training the department risks drawing the organisation's attention and potentially having to defend such expenditure. Never-the-less, value in business terms is described in pounds, dollars or other currency (Jackson, 1989) and the concept of value of training in such organisations is no different.

The utility of stakeholder analysis with respect to the practicality, type and scope of information gained in the empirical study therefore was limited to stakeholder group perceptions of the outcomes and effects of the programme. The stakeholder interviews were designed to gain other insights to the programme, most notably the content, structure and delivery approach of the programme and the perceived barriers to success.

Communication was via two reports; in 1993 and again in 1996 delivered to the programme steering committee. The 1993 report was delivered into an environment where management by metrics (see Chapter 8; section 8.3.4 Evaluation and the Management of Training) was the prevalent management ethos. Consequently, this had reinforced the steering committee's desire for counting the number and type of engineers trained in (attending) each module and the qualitative nature of reported stakeholder data was insufficiently persuasive to shift attention from counting. Although much of the data suggested changing the delivery approach to include suppliers, align the training to vehicle programme development, and train engineers at the time they needed to apply the training, evidence could also be found to support continuation of training by numbers. This in part was due to the steering committee's unwillingness to fully understand the nature of the data presented to them, but the large volume of data necessitated careful and time consuming study by the report's recipients. No actions resulted from the 1993 report.

The 1996 report was delivered in the final months of the programme, by which time it was too late to change the programme. As an assessment of the value of the programme to the organisation, the qualitative nature of the information did not facilitate value expression in financial terms and the report found both positive and negative aspects of the programme. In these senses, the report was inconclusive as an assessment of the value of the training to the organisation.

Within the literature of training, many writers (e.g. Deming, 1993; and Honey and Mumford, 1996) have suggested the concept of corporate learning whereby the organisation gains knowledge about its business by learning from corporate experiences. This type of learning was not observed to have occurred. Within months of the evaluation report, the director of Education and Training in Europe retired, responsibility of the programme was moved to Detroit where it was re-invented as a global technical training programme, and many of the programme staff moved to new areas of responsibility. The re-invented programme, which was a derivative of the Engineers' quality improvement training programme was developed in isolation of the reported evaluation data.

With respect to feasibility, the stakeholder analysis required cross cultural interviewing which was conducted by nationals of the respective countries. Using national interviewers is not only diplomatic in that the interviewers can communicate in the local language and observe local customs, it facilitates better understanding of interviewees. The process is however reliant on finding and training national interviewers and in this case where external interviewers were employed, the cost of using national interviewers is relatively high.

The interviewing process was concluded to be politically viable as all interviewees were observed to co-operate with the process and the organisation accepted the reported cultural differences as statement of fact and not criticism. This is considered to be due the level of management involvement in the planning and management of the interview process. Within Germany, where employee affairs are the subject of union/employer Works Council, careful consultation was undertaken in the planning stages to ensure all parties were committed to the process. This also ensured that employee rights were respected.

The stakeholder analysis was conducted with attention to ethical conduct and legal requirements. Parallel processes were operated in the UK and Germany, and as previously mentioned, the rights of employees were strictly observed. Interviews were conducted in confidence and although verbatim information was reported, this was accredited to the relevant sample group and not to individuals. The interviewee list remained confidential to interviewers and selected members of the evaluation group. The reporting of data was balanced and data analysis was conducted using the code-book to provide objective data which was not biased by perceived political persuasions.

Concerning accuracy, the technical adequacy of the interview schedule was subject to review by members of the programme development team and delivery teams in UK and Germany and were concluded to be adequate in this respect. Reliability was determined by comparison of interview data from the two interviewees which had been coded from the code-book by a third member of the evaluation group. Coding was found to be consistent with a high degree of replication in the coded data and therefore reliable (Guba and Lincoln, 1989). Validity was determined by a panel of programme trainers who coded interview transcripts into categories of meaning independently of the code-book. From this analysis a high degree of consistency was found between trainer interpretation of the data and from that of the code-book.

In addition to confirming the validity of this approach, the evaluation study illustrated that interviewers become more sensitive to the subtle culture differences within their own country. This is reflected in the literature; Connor (1985) found that by undertaking multinational evaluations, the evaluators became more sensitive to cross cultural differences made them challenge their own assumptions about their own culture. Connor concluded that these experiences may well make us more effective domestic evaluators.

7.3.2.6 Purpose of the Evaluation

As stated in chapter 5, the purpose of the evaluation was to evaluate the effectiveness of the programme in terms of the programme's aims (Figure 7-I). The aims of the programme are centred on improving the engineering process through the application of quality skills which support all aspects of the engineering process. Effectiveness was defined in terms of Kirkpatrick's four levels of evaluation regarding; participant's reactions to the training (level 1); participant learning (level 2); changes in participant's attitudes and behaviour (level III); and programme effectiveness (level IV).

Programme Aims
The aims of the engineers' quality improvement training programme are to give engineers: • in-depth quality improvement skills appropriate to their function • an appreciation of what skills are appropriate to other engineers in different functions and some understanding of these skills • an understanding of their roles and responsibilities within the engineering process and how they can directly influence and improve the quality of the final product or service • skills which support all aspects of Total Quality Excellence • An appreciation of the linkages between the quality improvement skills

Figure 7-I: Programme Aims

By adopting Kirkpatrick as the organising framework of the evaluation therefore, effectiveness was assumed to be associated with training outcomes.

7.3.2.6.1 Fulfilment of Purpose

With regard to the extent to which the study fulfilled its purpose, it is necessary to analyse the evaluation information in terms of each of the stated aims of the programme.

The programme aims are stated in general terms and are therefore subject to a wide interpretation. However, for the purposes of this review, they provide a directional indication of the types of evaluation information necessary to measure value.

With respect to giving engineers "in-depth quality improvement skills appropriate to their function", considering this in terms of behaviour as a function of the programme, the evaluation study was inconclusive as the behavioural element was restricted to only part of the curriculum (people skills elements) and only a pilot study was conducted. However from the stakeholder interviews, the content of the programme received widespread acclaim from all stakeholder groups (chapter 6; section 6.2.5.5.2) and the technical content was considered to be appropriate for quality improvement within the company (chapter 6; section 6.2.5.5.1) suggesting the programme could meet this aim if skills were applied.

From the programme foundation module knowledge assessment questionnaire (level 2 evaluation), engineers increased their knowledge of both technical and people skills by factors of 3.3 and 2.3 respectively (chapter 6; section 6.2.2.1). As the programme foundation module is concerned with the relationship of the programme quality improvement skills with the engineering process (chapter 4; sections 4.1 and 4.3.1.1), these results suggest that engineers did gain "an appreciation of what skills are appropriate to other engineers in different functions and some understanding of these skills" and "an understanding of their roles and responsibilities within the engineering process and how they can directly influence and improve the quality of the final product or service". These results also suggest the programme improves engineers' understanding of "the linkages between the quality improvement skills".

Concerning "skills which support all aspects of total quality excellence", stakeholder interviews in 1993 and 1995 found that trainee interviewees were unable to describe the relationship between the programme and total quality excellence (chapter 6; section 6.2.5.5.2), however there was considerable difference in responses from all other stakeholder groups who perceived the programme to support total quality excellence (chapter 6; 6.2.5.1.3). This suggests the programme did not provide engineers with a clear understanding of total quality excellence and they were therefore unable to describe its relationship. Observations made for the first aim are also relevant with respect to this aim.

7.3.2.6.2 Utility of Purpose

The purpose of the evaluation study was to *prove* the outcomes in terms of Easterby-Smith (1994). Training evaluation as defined by Easterby-Smith (1994) serves the purposes of; proving; improving; learning; and controlling. The practicality of this information can be analysed in terms of its influence on programme decisions and the resultant actions.

Although a significant amount of programme redevelopment was undertaken during implementation, this work was observed to be done in isolation from evaluation data (Brittle, 1995). Each module was the responsibility of the relevant development mini-team which had representation from the trainer groups in UK and Germany. Rework was based largely on trainers' classroom observations which were supplemented by trainee spoken comments. The timing of evaluation reports (see section 7.3.2.7) was observed to contribute to the lack of use of evaluation data in redevelopment work.

Within the wider context of organisational training, there is widespread agreement among researchers that the ultimate aim of training is improvement of performance in terms of organisational goals. Mulder et al (1995) comment that effective performance has become critical throughout the global economy and training is aimed at realising learning within organisations which directly or indirectly improve the effective performance. Goldstein (1993) emphasises the role of training within organisations to increase productivity, improve quality, reduce cycle time, become more service oriented to wards customers and reduce costs. Despite this, many researchers have found corporate training to be disconnected from organisational goals. Phillips (1991) comments that training programmes are often not linked specifically to strategies, challenges, or problems within an organisation. A view shared by Krijger and Pol (1995) who comment that training programmes are often criticised as cultural islands isolated from the organisation.

7.3.2.7 ***Communication***

A recurring theme throughout the review of the implementation evaluation components is the timeliness of communicating information to the programme decision-makers. Alkin (1988) comments that the value of evaluation reports after decisions have been made are of equal worth as ammunition received after the battle is over or food for a person who is dead from starvation. This is particularly true within commercial organisations, where change is becoming a way of life and a constant flow of current information is required to provide knowledge to decision-makers (Micklethwaite and Wooldridge, 1997).

The data provided from the evaluation of changes in attitude and of programme effectiveness in particular were delivered using detailed reports prepared for the programme steering committee. The information was made available too late to influence the management of the programme and in this sense the information was of little use.

Conversely, evaluations methods applied as part of the training; participant reactions; and knowledge gain questionnaires provided direct and immediate feedback to the trainers. Knowledge gain results were subsequently reported to the programme steering committee by way of a periodic evaluation report. Trainers received direct real-time spoken feedback on the training process allowing them to act if appropriate and feasible. Feedback of knowledge gain as a function of a given module was received by the trainers from course to course.

Similarly, results gathered from the pilot programme evaluation instruments provided direct real-time feedback to the programme development team, who were the decision makers and action takers. In these senses, the evaluation methods became tools of the programme trainers / developers and not of the evaluators as attitude, behaviour and stakeholder interviews methods were.

The trainer administered evaluation instruments were also observed to be adopted by the trainers and subject to improvement. Evaluation used in this way was observed to be used to manage and improve the programme.

7.3.2.8 Kirkpatrick Framework

With regard to the utility of Kirkpatrick's model for training evaluation, the meta-evaluation draws on practicality of information to audiences and scope and type of information criteria (section 7.2.3.1).

The practicality of information to audiences can be considered in terms of each of the four levels. Levels 1 and 2 of the evaluation were primarily conducted as part of the training modules, levels 3 and 4 were conducted away from training. In the case of this evaluation, levels 1 and 2 were administered by the trainers as part of the training responsibility, whereas levels 3 and 4 were administered by contract evaluators. Sample sizes for levels 1 ad 2 were close to 100% generating vast amounts of information. The sample sizes for levels 3 and 4 however were significantly less than 1% of the trained population. Despite this, the cost of level 3 and 4 evaluation far exceeded that of levels 1 and 2.

The primary reason for this is that in the case of levels 1 and 2, the evaluation instruments were integral to the training process and required little additional effort to collect and analyse the data. Levels 3 and 4 took the form of a separate additional activity where respondents were sought out and observed or interviewed, all of which was incremental to training. This is reflected in the literature of the application of Kirkpatrick's levels. Levels 1 and 2 are more frequently undertaken than levels 3 or 4.

In the normal series of training events, most contact between the training organisation and trainees occurs prior to and during training. Very little occurs after training. Identification of needs; requests for training; issuing of training joining instructions are regular occurrences for most training courses. These are necessary contacts and communication between trainer and trainee for training to occur. Evaluation activities conducted during these stages can be more readily integrated into the training process. Post training evaluation activities necessitate additional effort and contact with trainees.

Focusing on outcomes therefore has practicality constraints causing additional burden to the training department and, as demonstrated earlier, little is gained from this activity. Information which was considered to be of value in Kirkpatrick's terms merely corroborated that which was already known. In this study, evaluation data served little purpose to either programme improvement or decision making.

With regard to the scope and type of information, the framework directs the evaluation in terms of outcomes of training; reactions to the programme by the trainees; learning of taught principles, facts and procedures, trainee's behaviour as a result of training and results of the training with respect to the organisation. By the nature of this framework, the evaluation effort is therefore exclusively focused on outcomes. Process elements of training, such as parameters of the training design and aspects of the learner are, by definition, beyond the scope of Kirkpatrick's framework.

This lack of process information, compounded by the persuasion of evaluation as an external activity (chapter 4; section 4.6) and the consequent time-lag in evaluation information feedback (sections; 7.3.2.3; 7.3.2.4; and 7.3.2.5), provided insufficient information for the purposes of managing the programme on a weekly or monthly basis. Training performance in terms of its process parameters could only be inferred from analysis of outcome information. As previously concluded, this information became available too late and would require further investigative activity. Within any organisation, investigation requires resources and the investigation process in itself is not value added; it is only the actions which result from an investigation which are likely to contribute to the business.

The predominance of Kirkpatrick in the literature relating to training evaluation in a commercial context is a confining and therefore limiting factor and history has shown that evaluation at level 4 is an illusive activity for training practitioners and researchers alike (chapter 2; section 2.4.4).

7.4 Conclusions

With respect to the role of training evaluation in the improvement of training in a commercial context (focusing question 1: chapter 2: section 2.6.1), from the meta-evaluation three general conclusions can be drawn which have wider implications for training evaluation with respect to (i) ownership of evaluation; (ii) the utility of Kirkpatrick's model for training evaluation; and (iii) value judgement in a commercial organisational context.

7.4.1 Ownership of the Evaluation

A predominant feature of the empirical study concerns ownership of the evaluation. In this context, ownership is a role concept with respect to the instruments of the evaluation and the information provided. Behavioural, attitude and results evaluation instruments were all employed by members of the evaluation team, with no involvement of the delivery team in the implementation (data collection and analysis) stage. This approach to evaluation is a recurring theme in the literature. Many writers conceive evaluation in terms of evaluator - client relationships; Easterby-Smith (1994) refers to impartiality issues resulting from close relationships between evaluators and clients; Reid and Barrington (1994) advocate that the overall responsibility for evaluation is best vested in a neutral party, such as an external consultant, to ensure impartiality; and Bramley

(1991) refers to issues of reporting faced by an evaluator with respect to the interests of clients. The inference to be drawn from the literature is that training evaluation is a set of practices conducted by an evaluator to meet the purposes of the client. The tools of evaluation; data collection; data analysis; and communication, are those of the evaluator.

A further inference to be drawn from the literature in this respect is the nature and frequency of the flow of evaluation information. The writers cited refer to evaluation reports which provided results and conclusions. In some cases these are formative evaluation reports and in others summative evaluation reports. In either case, evaluation is conceived as an event or series of events culminating in an evaluation report.

Finally, the notion of evaluator - client relationships holds to the view that evaluation is done to someone or something. The evaluator examines a training programme; its learners and instructors, using a range of methods to collect information sought by the client. The information is analysed, collated and conclusions are drawn which are presented back to the client. Associated with this is the belief that evaluation is a cost added activity. Many writers (i.e. Rossi and Freeman, 1993;) have commented on the cost of evaluation and the need for budgeting. This reflects their perception of evaluation as an added, as opposed integral activity of an evaluand.

The emergence of the notion of programme evaluation stakeholders (Goldstein, 1986; Patrick, 1992; Basarab and Root), where learners and instructors are identified as stakeholders, has done little to shift the underlying assumption that evaluation is the discipline of the evaluator. The predominance of this paradigm in the literature has significantly influenced the way evaluation is conceived and the nature of its implementation.

7.4.2 Utility of Kirkpatrick's Framework

The evaluation was conceptualised using Kirkpatrick as an organising framework. The emphasis was therefore on outcomes conceived in terms of; reactions; knowledge gain; changes in attitude and behaviour; and organisational perception of results.

The role and utility of the evaluation with respect to the management of training is central to training evaluation within the context of commercial organisations. The function of the training department is to provide the organisation's human resources with the knowledge, attitudes and skills to perform their jobs (Walker, 1992) and training, therefore, offers a means of realising business success (Moore and Seidner, 1998). The process of training is a cost to the organisation and the role of training management is one of reducing costs and increasing learning. Both cost reduction and increase in learning is achieved through continuous improvement of training.

Continuous improvement is the primary goal of management (Deming, 1986) and this holds for the management of training. Information which does not support this goal becomes secondary to information that does. Evaluation in the sense that it is applied in this study does not directly support this goal and therefore contributes more to the costs of training than to its improvement.

Reactions and knowledge gain had been considered as part of the development of the programme and these were concluded to be satisfactory in terms of the expectations of the development team. During the initial stages of full implementation (first six months) the conclusions drawn from the pilot study were confirmed and the results from reactions checks and knowledge measures were stable and predictable. The attention to trainer selection and development (chapter 4; section 4.5.3.1; and section 4.5.3.2) assured trainer consistency and although trainers working on the programme changed during the evaluation period, participant reactions and knowledge gain remained stable. In this sense, the training design and materials were observed to be robust to trainer and trainee variation.

Attitude measures and behavioural observation were resource intensive particularly with regard to the latter where only limited observational data was collected. The attitude measure questionnaire was administered using the company's internal electronic mail system, providing instant access to all staff employees (NB with the company of interest all engineering personnel targeted by the programme are staff employees). Whilst this communication system facilitated the administration of the evaluation questionnaire, the easy access to e-mail has resulted in a plethora of questionnaire surveys on a wide range of topics within the company concerned. This has been observed to result in a general disregard to surveys by staff employees and lower rates of return (FMC, 1997).

The measure of perceptions of the organisation through stakeholder interviews revealed some interesting insights into the perception of effectiveness of the programme at different levels of the organisation. Attempts to elicit further information from this data using attributional analysis provided little information on the critical success factors of the programme. It largely confirmed feedback which had previously been received from trainer-trainee dialogue during the programme delivery and through post-training consulting projects. Perceptions of senior managers had been elicited from those managers who were part of the programme steering committee. In this sense the stakeholder interview process revealed little which was not previously known.

Evaluation as a measure of outcomes considers only part, albeit an important part, of training. Kirkpatrick's organising framework takes no direct account of the many process parameters of learning and training, i.e.; aspects of the learner (chapter 1; section 1.2.5); and conditions of training transfer (chapter 1; section 1.3.2). This reflects the widely held view within the literature of evaluation as value judgement (chapter 2; section 2.4.5), where value is associated solely with programme outcomes.

7.4.3 Value Judgement in a Commercial Context

The role of training within commercial organisations is well documented and there is a considerable degree of agreement between writers. Brinkerhoff (1987) comments that HRD (training) approaches improve individual and organisational performance through learning. Sloman (1994) emphasises the importance of a training strategy which is directly aligned to the organisation's business strategy (goals and processes). Phillips (1991) reflects this persuasion of training as a strategy of improving business performance by describing training as results-oriented and bottomline contributing. The role of training in modern commercial organisations is one of developing knowledge, attitudes, and behaviours which directly or indirectly improve commercial performance.

Evaluation writers have broadly interpreted evaluation in this context as one of providing feedback to training stakeholders on the value of the training in terms of agreed value criteria (i.e. Newby, 1992; Easterby-Smith, 1994; Kirkpatrick, 1994; and Moore and Sedner, 1998). Kirkpatrick's framework provides a range of value criteria in this respect; trainee reactions; trainee learning (knowledge gain); trainee behaviours; and organisational results. Trainers and training managers are given feedback on whether trainees reacted favourably or otherwise to the training; and identified training stakeholders are provided with value judgements with respect to learning, behaviour, and results.

Two key issues emerge from the empirical study within the value judgement paradigm; (i) the utility of value judgement information; and (ii) the roles and relationships of training operators in the evaluation process. Whilst many writers agree that evaluation information is of little use if it is not used for decision making or programme improvement, exponents argue a macro-level feedback loop. Formative or summative evaluation data is used to measure performance against aims or intended outcomes which lead to corrective actions. Whilst this type of feedback provides for redesigns to existing programmes or inputs to future programmes, it is incapable of facilitating ongoing or continuous improvement; improvement actions made on a daily or weekly basis by all levels of people within the training programme.

The evaluation featured in part II of this thesis was initiated and managed by the organisation's European central training staffs. All decisions relating to evaluation policy, methodology and data reporting ultimately lay with the central staff. Furthermore, the comparison of evaluation data between national companies, and in particular between the UK and German reinforced the policing role of the central staff.

The history of the programme's conception and development had formed the pattern of roles and relationships between the various components of the organisation long before the introduction of the evaluation. Political battles had been fought and won by the central staffs with its national partners with regard to the design of training materials, the inclusion of content, the selection and training of trainers. Whilst this in itself had not established the relationships between the components, it had reinforced the nature of those relationships and the distrust which existed between the central staff and its national allies.

The evaluation of the programme had largely been conceptualised as an evaluation by the central staff of the national companies programme delivery operations. The evaluator-client relationship, implicit in much of the literature (chapter 4; section 4.6), was a critical feature in the empirical study and reinforced the central staff - national company relationship with respect to the evaluation. The levels of co-operation by those involved in different aspects of the training are symptomatic of this evaluator-client relationship.

Evaluation is therefore firmly couched in a value-judgement paradigm, where the role of evaluation is to provide conclusions on the value of training in terms of agreed criteria. Kirkpatrick (1994) introduces evaluation as "the reason for evaluating is to determine the effectiveness of a training programme. When the evaluation is done, we can hope that the results are positive and gratifying, both for those responsible for the programme and for upper-level managers who will make decisions based on their evaluation of the programme" (p.3).

Accepting the notion of a value judgement paradigm and its implications on research and practice, a restructuring is necessary which facilitates its integration with the theories and practice of learning and training (research focusing question 2: chapter 2: section 2.6.1). The next chapter begins to explore such a restructuring.

Chapter 8

Conclusion: Directions for Practice and Research

8. Conclusion: Directions for Practice and Research

> *As in manufacture so in science - retooling is an extravagance to be reserved for the occasion that demands. The significance of crises is the indication they provide that an occasion for retooling has arrived.*
>
> *- Thomas S. Kuhn (1962)*

8.1 Introduction

This chapter is based on the conclusions to chapter 7 and on inferences drawn from the literature with respect to learning; training, evaluation, and measurement. From the reviews of the literature, it may be inferred that the evaluation of training is in a state of crisis; the lack of reported evaluation studies in terms of Kirkpatrick, contrasted with its popularity among writers of training evaluation is a cause for concern and in this sense, the evaluation of training is in a state of crisis. This view is confirmed by the conclusions drawn in chapter 7, where the evaluation of the training programme was found to be of little utility to the organisation. A re-tooling or paradigm shift is necessary and this chapter is intended to provide a directional framework for training evaluation in commercial organisations.

As previously discussed (chapter 2), Kirkpatrick's framework has been in existence since 1960 and has been elaborated upon, most notably by Hamblin (1974). Its popularity in training has resulted in many and often unsuccessful attempts to evaluate training in these terms (Alliger and Janak, 1989). The empirical study identified several weaknesses with this approach, particularly with respect to its utility for the management of training; a stated aim of Kirkpatrick (1994). The persuasion to evaluation as an activity external to the training programme in the literature (i.e. Madaus et al, 1996; Guba and Lincoln, 1989) has given rise to the practice of using evaluation as a tool of the evaluator to evaluate others and this approach was largely adopted in the empirical study. The meta-evaluation provides evidence of problems with ownership and lack of co-operation from those involved in the programme and by the nature of this approach the decision making information generated by the evaluation only became available long after decisions had been taken. In this sense, the information was of little, if any, utility value to the organisation.

8.2 Value Judgement Paradigm

The conclusions to be drawn from the literature with respect to the evaluation of training is that is it largely held within a paradigm of judgement of value. The emphasis on judgement of worth has held training evaluation in a relative static position in comparison with developments in other areas of science and research. The popularity of Kirkpatrick has continued to reinforce this paradigm amongst training practitioners; it has largely defined the language of evaluation with professional journals referring to level 1; level 2; etc. evaluation without the need to explain the reference or origin. For the evaluation of training, practitioners are encouraged to think in terms of Kirkpatrick (e.g. Basarab and Root, 1994) and this was the case with the empirical study.

For the purposes of this analysis, I refer to this widely held belief system as a value judgement paradigm. Value judgement refers to the assessment of achievement of the training programme; and paradigm in the sense of Kuhn (1970) as the constellation of beliefs, values, rules and techniques shared by members of a community. The significance of this paradigm to the evaluation of training is its defining nature of what evaluation is; attention to outcomes and the exclusion of other parameters from the study of learning and training.

8.2.1 Evaluation: Investigation or feedback

Within this value judgement paradigm, the underlying fundamental question of evaluation is one of investigation; 'how well did we do?", as opposed to "how are we doing?". The former points to a strategy of investigation - a one-off type study which is distinguishable from its evaluand. The latter points towards a continuous process of getting performance information as part of a feedback loop and is integral and indistinguishable from the evaluand.

Kirkpatrick (1994) asserts that the purpose of evaluation is to; (i) justify the existence of the training department by showing how it contributes to the organisation's objectives and goals; (ii) decide whether to continue or discontinue training programmes; and (iii) to gain information on how to improve future training programmes. This view of evaluation is endorsed by Bramley (1997) who comments that training evaluation is the process of establishing the worth of the training. Kirkpatrick and Bramley are not alone in their 'how well did we do?' conceptualisation of evaluation and this is implicit in much of the literature on the subject.

From the empirical study meta-evaluation (chapter 7; section 7.3.2.7), it was concluded that the programme trainers, developers and managers required real-time 'how are we doing?' type information in order to make daily decisions about the programme and continuously improve the training delivery and administration in support of programme and organisational goals. The outcome nature of the evaluation information and the time-lag inherent in the evaluation communication process between evaluator and client yielded information with little utility to these groups.

8.2.2 Characteristics of a Value Judgement Paradigm

Training evaluation within this value judgement paradigm is characterised by; (i) outcome oriented information; (ii) exclusion of other learning / training parameters; (iii) evaluator-client relationships; and (iv) lapsed-time feedback of information. The outcome orientation of Kirkpatrick directs the evaluator and shapes evaluation in these terms. Whilst efforts are made to provide information for programme improvement, this is a secondary action of the evaluation which requires further investigation in some form.

The exclusion of other learning and training parameters (factors) as a direct measure of the training evaluation is, in the view of the writer, a major deficiency of the empirical study. From reviews of the literature in chapter 1, a vast body of knowledge exists on the factors which affect learners and training process. Most notably in the area of transfer of training which is an essential element of training which is intended to support organisational goals (chapter 7; section 7.4.3). Learning and training is the subject of continuing research and yet little attention is paid to this by training evaluation practitioners; observations made by Baldwin and Ford (1988) that training research and practice are largely divorced still hold true today.

Evaluator - client relationships (chapter 7; section 7.4.1) place evaluation as the tool of the evaluator to collect and analyse data to provide information for the client to make value judgements of the programme. Lapsed time feedback was observed to be a feature of the empirical study and in this respect was concluded to have a negative impact on the utility of the evaluation information.

Training evaluation, as widely conceptualised within the literature, forms part of the final stages of the training process (chapter 1; section 1.3.3). In this sense training evaluation can be described as a down-stream activity, where the training process is an upstream to downstream series of activities. The significance of this downstream nature of evaluation will be made clearer to the reader in the next section.

8.3 Conceptualisation of Evaluation in Commerce and Industry

Evaluation is widely associated with education, training and other social programmes and policies. The purposes of evaluation given in the literature are broad ranging, but can be summarised as *proving, improving, learning* and *controlling* (Easterby-Smith, 1994).

Within commercial and industrial contexts, activities of which the purposes *are proving, improving, learning* and *controlling* are collectively described as quality control, or more recently total quality management (Dale, 1994). The goals and activities of quality control broadly parallel those of evaluation. The notion of quality is essentially a value judgement of a product or service and the goal of quality control is to assure products and services (outputs) which are of high quality (high value) in terms of an identified criteria.

Given the dominance of Kirkpatrick and the emphasis on measuring outcomes in training evaluation practice, the origins and evolution of quality control has particular relevance to this study.

8.3.1 Evolution of Quality Management

Dale et al (1994) suggest that quality management has evolved through 4 stages; (i) inspection; (ii) quality control; (iii) quality assurance; and (iv) total quality management.

Inspection was concerned with the examination of a product or service using measurement or testing and comparison to specified requirements. The responsibility for inspection was with an inspector who was not part of the organisation responsible for making the product or providing the service. Information from the inspection activity was primarily for deciding whether a product or service was acceptable or unacceptable. Whilst inspection information was used to improve products and services, it was in the sense of problem solving or correction and the activities were not integrated into the making of the product or delivery of the service.

Taylor (1919) identified inspection as a specific task in his work on scientific management commenting 'the inspector is responsible for the quality of the work, and both the workmen and the speed bosses (who see that proper cutting tools are used, that the work is properly driven, and that cuts are started in the right part of the piece) must see that the work is finished to suit him'. This inspection role was reactive and focused on the outcomes of the process.

Quality control evolved from, and incorporated many of the concepts of inspection. The emphasis remained with the assessment of outputs, however attention was paid to the control of incoming raw materials and other activities which contributed to the making of the product. This was largely done by an incoming inspection activity, operated in a similar way to that of the outputs. The responsibility for quality and quality control activities remained with the Quality Control (QC) department, who were independent of those making the product or delivering the service. The independence was considered to be important for ensuring the impartial objectivity of the QC department.

Quality assurance evolved out of the recognition that the 'find and fix' activities of inspection and quality control did not prevent problems from occurring, they merely enabled errors to be found and either corrected or deleted. Quality assurance was largely concerned with prevention and activities were applied which to facilitate the prevention of production problems. Deming (1982), based on the work of Shewhart (1931), emphasised the importance of data and the use of (Shewhart) control charts to draw inferences about the production run from data collected on a sample of products. Control charts allowed for assignable (special) causes to be distinguished from natural inherent (common) variation in the process. The former represent problems that may be addressed by the production operators taking problem solving actions; the latter are inherent in the process and require management action.

Another aspect was the control of the design of a product where problem prevention techniques, such as Failure Mode and Effects Analysis (Johnson, 1997), were used to identify potential problem areas in a process or product and eliminate their causes from a design. This defect detection philosophy to quality management resulted from an increasing recognition for the need to prevent problems occurring and the associated costs of scrap and dissatisfied customers.

Total quality management, refined from Feigenbaum's (1961) original concept of total quality control, involves the application of quality methods to all aspects of the business including customers and suppliers. The emphasis is on the customer, with quality being redefined as 'meeting the needs and expectations of the customer' (Henshall, 1992). Quality is the responsibility of those who operate the process and activities are directed at those factors which are known, or believed, to affect quality. Another important aspect of total quality management is Deming's (1982) contention that quality management activities should serve to continually provide intimate knowledge of the process and product. Deming's notion of 'intimate knowledge' refers to very specific and fine detail understanding of the performance of machines, materials, methods, people and process environment by those who operate the process on a daily basis, as well as those who manage the process.

A further significant evolution of quality was the introduction of the notion of positive and negative quality. Kano and Takashi (1979), transferred and applied concepts from Herzberg's two factor theory of worker satisfaction (Hertzberg et al, 1959) to quality in customer terms. They characterised quality by two types of factors; (i) negative quality factors which when absent dissatisfied customers, but when present did not cause customer satisfaction; and (ii) positive quality factors which when present caused customer satisfaction, but when absent did not dissatisfy customers. The notion of positive quality therefore is applied, where quality is conceived in customer terms and is not limited to meeting specification or the elimination of problems, but in exceeding the needs of the customer (Bossert, 1991).

Crosby (1979, 1984), Dale (1989), Deming (1982, 1983), Juran (1992), Oakland (1993), Peters and Waterman (1982), and Zairi (1991) have proposed models and methods to achieve continuous improvement and the involvement of the entire workforce to focus on the satisfaction of the customer, both internal and external. The range of activities has therefore widened and incorporates aspects such as teamwork, creativity and innovation, factors which previously were perceived to be beyond the scope of quality management activities. The management of quality is integrated into the overall design, planning and implementation of the manufacture of products or the provision of services.

Unfortunately the concept of total quality management has largely been misunderstood and adopted a faddish status in many organisations (Calvert et al, 1994). The evolution to total quality management has shifted paradigms of quality from measure and control in terms of a specification to customer focused and continuous improvement. An essential characteristic of quality management, with respect to continuous improvement, has been the real-time nature of data collection and analysis instruments, allowing appropriate stakeholders in the design and manufacture of products to react accordingly. The second major shift has been the move away from independent inspectors to becoming the role of those making or providing the service. Twenty years ago, the notion that a production operator could be responsible for controlling his/her own quality would have been regarded as absurd, with critics arguing that independence and impartiality were central to assessing and controlling quality. These shifts have largely been overlooked in many non-manufacturing sectors where attempts have been made to implement total quality management (Dale et al, 1994).

8.3.2 Parallels with Evaluation of Training

Parallels can be drawn from the literature relating to the evaluation of training and the management of quality. Juran (1988, 1989) describes quality management in terms of planning, control and improvement and offers a 'road-map' for quality management. The road-map consists of nine steps which describes the management of quality as an integral aspect of the overall planning and implementation of a product or service (Figure 8-I).

Stage Number	Road-Map Stage
1	Identify who are the customers
2	Determine the needs of those customers
3	Translate those needs into our language
4	Develop a product that can respond to those needs
5	Optimise the product features so as to meet our needs as well as customer needs
6	Develop a process which is able to produce the product
7	Optimise the process
8	Prove that the process can produce the product under operating conditions
9	Transfer the process to operations

Figure 8-I: Juran's Quality Management Road-Map

Juran's road-map mirrors the training and evaluation processes of Camp et al (1986); Jackson (1989); Bramley (1991); and Newby (1992). The reader will recall from chapters 1 and 2 that these may be generalised into the process given in Figure 8-II.

Key Activities	Stage Number	Process Stage
Identification of training needs	1	Organisational Analysis
	2	Task Analysis
	3	Persons Analysis
Training Design and Development	4	Write Training Objectives
	5	Design Curriculum
	6	Develop training content and methods
Training Delivery	7	Plan logistics
	8	Deliver training
Evaluate Training	9	Collect and analyse evaluation data
	10	Communicate training results

Figure 8-II: Generalised Training and Evaluation Process

Stages 1-3 of Juran's road-map correlate to stages 1-4 of the generalised training process. These are both concerned with the identification of needs and the definition of the target audience. Juran's step 3; translation of needs into our language, refers to the specification of product in engineering terms (for the purposes of design) and are similar to the establishment of training objectives for the purposes of training design. Stages 4-6 of Juran's road-map correlate to stages 5-6 in that they are concerned with the development of product and process to meet the specified requirements. Stages 8 and 9 of Juran's road map correlate to stages 7-10 of the generalised model of training as they are both concerned with execution and evaluation.

Drawing this parallel is not new as several writers have contrasted evaluation with quality control; Thackwray (1997) draws on definitions of quality from Juran (1988), Crosby (1984), Taguchi (1985), and Feigenbaum (1986) and, by arguing that these can be translated into an educational context, suggests that evaluation is the significant contributor to quality control; Bramley (1997) suggests the most common reason for evaluating training is to provide quality control over the design and delivery of training activities (pp 5-6). Basarab (1998) emphasises customer focus and continuous improvement as essential elements of training evaluation under the heading of 'total customer satisfaction'. Whilst Basarab has adopted many of the techniques of quality management, these efforts continue to be held in a value judgement paradigm; aspects such as real-time feedback, evaluation integral to the training process, development of intimate knowledge of the training process through the measurement and study of its parameters are not features of this interpretation.

8.3.3 Discovery Process

The evolution of contemporary quality management principles has evolved through four stages as identified by Dale et al (1994). Each stage has evolved from the previous, as knowledge is gained through experience. Evolutionary processes of this nature are time consuming. The economic pressures which exist in the commercial world have acted as the primary driver to this discovery process. Sadly, or thankfully, such pressures do not exist to the same degree in the training world, although the trend toward emphasising the need for bottomline results from training (i.e. Walker, 1992; Phillips, 1994, 1998; Mulder et al, 1995; and Brown, 1998) suggest this pressure may be increasing.

Given the similarities between inspection and the general contemporary evaluation; measurement of outputs; little regard to process inputs; evaluator - client relationship; and time-lapsed feedback of information, the development of total quality management offers opportunities to the training community to redefine its thinking.

The inspection and quality control stages of the management of quality were primarily concerned with outputs and were largely divorced from the activity they attempted to evaluate. They were integrated in the sense that they assessed a product or service against a specification, in a similar way to evaluating training in terms of its specified goals or objectives, but the underlying philosophy was one of independence from the evaluand. Inspectors, like evaluators, were necessarily positioned independently from the operators of the process to give them the independence which was considered necessary for them to make objective observations and decisions about the evaluand.

8.3.3.1 Operator Autonomy

The shift towards operator autonomy with respect to quality management is significant. The concept of quality circles, features largely in the literature as a means of facilitating this autonomy and collective decision making which adds value (or quality) to the product or service. Dr Kaoru Ishikawa, the noted authority on quality circles (Nemoto, 1987) describes them as a group of workers doing similar work who voluntarily and regularly meet in normal working time to identify analyse and solve work related problems and to recommend solutions to management (Ishikawa, 1985).

Munro-Faure and Munro Faure (1992) emphasise the importance of training members of quality circles, or as they prefer to term them 'progress groups', in methods of quality improvement and teamwork as well as providing members with an 'ongoing education' of the important quality related issues within their local context and with respect to the company's product or service performance in the market place. Munro-Faure and Munro Faure suggest that quality circles are the essential element to effective quality management because they facilitate operator or worker knowledge in the management process and empower workers to take action without real-time inhibiting bureaucracy - a recurring feature identified in the empirical study.

From Nemoto's (1987) treatment of quality circles, their purpose can be considered to be two fold; (i) they facilitate the identification, analysis and resolution of quality concerns; and (ii) they serve as a means of communicating the importance of quality and placing emphasis on contemporary thinking and issues to operators and supervisors.

Hill (1994) notes that in their peak (mid-1980s), quality circles could be found in at least 400 British companies, making this numerically the largest innovation in participative quality management. By the end of the decade, however, most companies had wound up their programmes despite the importance attached to quality improvement and quality circles by senior management. Hill's analysis of this failure concludes that unless quality circles are integrated into the normal operation or activity of a company they are unlikely to work.

8.3.4 Evaluation and the Management of Training

The role of evaluation in the management of training has largely been neglected in the literature, despite many writer's claims that a purpose of evaluation is improvement (i.e. Easterby-Smith, 1994; Morrison, 1993; Basarab and Root, 1994). Sloman (1994) cites the Training Agency's (1989) National Training Survey and conclude that 'the number of training days received per annum remains the best indicator of the health of training in an organisation'. The management of the programme subject to empirical study was consistent with National Training Survey's conclusion; reported training attendance figures was a permanent feature of the Programme Steering Committee meetings with each national company reporting the numbers of engineers trained in each module. No attention was given to evaluation measures of training performance.

The reasons for this are two-fold; (i) the evaluation information provided was on a bi-annual frequency and therefore not available when required; (ii) the type of information did not directly relate the manageable training attributes. Outcomes measures provide an indication of the way things are and not why. Causes (controllable or noise factors of the training process) of success or failure were not subject to direct measures.

The utility, and therefore value, or evaluation information is determined by the way it is used in the management of the programme. Evaluation information must necessarily be provided when it is required and in a form which can be directly translated into action.

8.4 Maximising Value Paradigm

The literature on training evaluation has traditionally focused on the development of strategies for measuring training effectiveness in pursuit of operational goals. More recently greater emphasis has been placed on measuring return on investment, contrasting the costs of training with its contribution to an organisation's bottom-line profit/loss (i.e.; Walker, 1992; Phillips, 1994, 1998; and Mulder et al, 1995).

From the literature, or lack of it, of successful and convincing applications of this orientation of training evaluation, a reassessment and re-scoping of evaluation is desperately required. This is born out by the sterility of the empirical study. Such reassessment dictates the need for a wider view of the conceptualisation of evaluation, not just in training, education and social programmes, but in the wider context of management, industry and commerce.

Brinkerhoff (1995) comments that the purpose of training is to add value to the organisation and as such training practitioners should be concerned with the instrumental value of training. In this context, the term instrumental value refers to the contribution of training to the organisation's goals.

8.4.1 Beliefs, Values and Basic Assumptions

Beliefs and values provide the fundamental structure of any paradigm; they reflect the unspoken basic assumptions which are shared by a community and with which members of the community interpret the world. A belief system underpinning a maximising value paradigm is characterised by; training evaluation as an integrated part of learning and training; continuously gaining knowledge about the learning and training process in a specific context through evaluation; and continuously improving the learning and training process on current and future interventions.

A maximising value paradigm, therefore, is characterised by; learner measurement of training outcomes; management of (learning and training) process parameters; real-time feedback; and operator (trainer and learner) control.

8.4.2 Redefining Training Evaluation

Definitions are critical to a field of study as they provide a statement of the nature, properties, scope, or essential qualities of an entity (chapter 1; section 1.2.1). Research students are encouraged to consider definitions of their selected topic (Howard and Sharp, 1983; Cooper, 1984) and definitions of subject concepts are a feature of literature works, often referenced in subject indexes. Definitions are a powerful shaping influence on readers of a given topic.

Training evaluation is concerned with training and evaluation and, as learning is central to the training process, its definition must draw on these three fields of study. Training evaluation therefore becomes an integration of (i) learning; (ii) training; and (ii) evaluation, as opposed to being defined as a distinct discipline.

Definitions of learning (chapter 1; section 1.2.1) summarily describe the phenomena which is made up of interrelated constructs (concepts), definitions, and propositions that present a systematic view. These theories provide the general explanations of learning and three theories are reviewed in Part I of this study (chapter 1; sections 1.2.2.1; 1.2.2.2; and 1.2.2.3). The internal mechanisms of the learner are considered, as are the external factors to the learner. Learning is subject to a range of criteria, of which training (instruction) forms just part. Aspects such as learner motivation, ability, and perceptions of the work environment (chapter 1; section 1.2.5) are three selected internal criteria featured in the literature.

Within a maximising value paradigm of evaluation, training evaluation might therefore be defined as:

> *"a management philosophy embracing all activities of learning and training through which identified goals of the organisation are satisfied in the most efficient and cost effective way by maximising the potential of all those employed in the training process in a continuing drive for improvement"*

The definition is offered as it characterises training evaluation in a broader sense than that of Tyler (1950) or Patrick (1992) and extends from the definitions offered by Warr et al (1978) and Goldstein (1986).

The definition comprises five key statements; management philosophy; all activities of learning and training; identified goals; all those employed; and continuing drive for improvement.

Management philosophy refers to the values and goals which are actively encouraged; the establishment of a vision of success and how that vision should be achieved in the sense, for example, of Peters (1988), or of Petersen and Hillkirk (1991), or of Deming (1993). All activities of learning and training embrace the vast array of current knowledge with respect to learning and training, for example; individual and organisational learning; characteristics of learners and of learning; training or learning intervention design; and the transfer of training and the transfer environment. Identified goals refer to the targets or objectives in the established sense of Kirkpatrick, Phillips and others.

Evaluation in this sense provides the methods of data collection, analysis and communication appropriate to all those employed in learning and training; the stakeholders in the process, i.e. learners, trainers, and training managers. Continuing drive for improvement is a state of mind; a conditioned desire of everyone involved in learning and training.

8.4.3 A General Framework for Evaluating Training

If one accepts the need for a shift to a maximising value paradigm, and is in general agreement with the underlying principles which have been offered, the foundations for a new approach to training can be explored. This section is not intended to define a set of evaluation procedures, but to provide a general guiding framework for development through future practice and research.

Within a maximising value paradigm, training evaluation becomes integral to the process of continuous improvement. It provides information for decision-making at all levels of the training or learning intervention. Furthermore, the provision of the information is in a form that facilitates measurable actions which are within the control of those for whom it is intended and it is provided when it is needed; is of utility value to the recipient, and therefore the organisation.

8.4.3.1 Learning and Training as Process

A common notion among writers of TQM is defining, or modelling, any activity as process. The concept of modelling is to selectively represent the important features of a phenomenon in the simplest form which meets the intended needs (chapter 3; section 3.2.1) and so process modelling, therefore, is distinguishing the important parameters (i.e. inputs, outputs, resources, and controls) of learning and training. Such parameters can be derived from; the vast body of literature; the particular design of the learning intervention; and from past experiences of learning and training within an organisation.

From the reviews of the literature in chapter 1, and from the empirical study (chapter 4), training is a multi-faceted activity comprising of several inter-related set of serial and parallel processes and sub-processes. The primary processes of the programme studied are given in Table 8-I.

Primary Process	Process Elements
Training Development	Needs Analysis Training Objectives Instructional Design (see Wager et al - Principles of instructional design and Reigeluth - Instructional design theories and models (Reigeluth, 1983) Instructional development
Training Delivery	Training participation Training outcomes (Kirkpatrick's levels of evaluation) Post training course activities (on the part of the trainee and the trainer) Delivery management
Training Attendance	Identifying participants Preparing participants Attitude and role of the manager or supervisor Attitude and role of colleagues Status of adoption of/ resistance to programme concepts in the area
Trainer selection and Development	Selection Trainer training trainer performance - in and out the training room
Applications Consultancy	Getting in Getting on Getting out outcomes

Table 8-I: Primary Process Elements of the Engineers' Quality Improvement Training Programme

Each primary process is comprised of multiple process elements. For each element, a process model can be derived which selectively represents the important features of learning and training which can be the subjects of the evaluation.

8.4.3.2 Managing Training Parameters

From the simple model of the training process, inputs to the process are the obvious parameters for consideration. Using the programme subject of the empirical study, trainees, trainers and materials are some of the the primary inputs to the process. Prior to and during training, contact between training administrators and trainers and the trainees is part of the normal business. Any evaluation activity undertaken during these stages can be easily integrated with the normal training process, unlike post training contact which, from the empirical study, requires additional resource. An observation of note in this respect, made during the implementation of the programme (Brittle, 1995), was attendance by trainees who were not in a state of readiness to apply the skills. As the programme in question related to the engineering process, training in the respective disciplines was more likely to be temporally relevant immediately prior to engineers having to apply those skills.

Persons analysis (chapter 1; section 1.3.4.3), whilst regarded as providing the means of identifying trainees who are in a state of readiness to learn (Wexley, 1984; Reid and Barrington, 1994), is not actively applied by training practitioners (Tannenbaum and Yukl, 1992). Measurement of readiness to learning, and the subsequent use of this information in the management of the programme, offers training managers the information by which to manage the input process.

8.4.3.3 Learner Ownership of Outcomes: Self Evaluation and Reporting

From the reviews of the literature relating to; learning; conceptualisations of evaluation; and evaluation methodology, and from the empirical study, there is considerable evidence in support of self evaluation.

The need for training outcome data, particularly data which relates to 'bottom-line impact (i.e. Walker, 1992; Phillips, 1994; and Mulder et al, 1995), is widely considered to be of most importance to training evaluation. From the management of quality literature, the shift to operator evaluation of the process has not undermined the integrity of the measurement data as critics had argued it would in the early stages of the evolution of total quality management. The resultant effect was greater ownership of the process, a significant reduction in quality control costs, and real-time feedback of process performance. (Ward and Dale, 1994).

From the empirical study, with support from the literature, measurement of behaviours and results from training is illusive. The main inhibitors to this type of evaluation appear to be: (i) the practical constraints in collecting the data; (ii) the costs of collecting the data; and (iii) isolation of training effects from other factors. The practical constraints are largely concerned with the time delay after training and trainees no longer being in contact with the training department. Whereas immediate reactions to training (Kirkpatrick level 1) evaluation data and knowledge gain

(Kirkpatrick level 2 data) can be collected immediately before trainees depart from the training environment, behaviours and results cannot be evaluated until the training has returned to their job and transferred the training. If behaviours are to be used to imply learning, then sufficient time lapse must occur to infer the characteristic of permanency.

The costs associated with this type of data collection are in the administering of evaluation instruments by a third party evaluator, or even where the evaluation is undertaken by the training department staff, additional costs are incurred in seeking out trainees, and collecting and analysing data at a time when the trainees are using the skill. In the case of the empirical study, only a small sample was affordable as the costs of data collection were relatively high compared to the direct training cost.

With regard to isolating the effects due to training from those due to other factors, experimental / control group studies are expensive. It is also very difficult, if not impossible, to identify a control group which is unaffected by organisational learning gain caused by the influences of those who have attended training. Learners are more likely to predict the effectiveness of their own training with respect to the benefits to the organisation, than an unfamiliar external evaluator.

Drawing on Bandura et al, 1977 (chapter 1: section 1.2.2.3: Social Learning Theory), and within the learning cycle of Kolb, 1984 (chapter 1: section 1.2.3.1), the learner as self evaluator is positioned, if sufficiently sensitised, to take account of learning resulting from his or her environment as part of the post training learning cycle (Figure 8-III). These influences, whether positive or negative with respect to the aim of the training, are important factors in the training process.

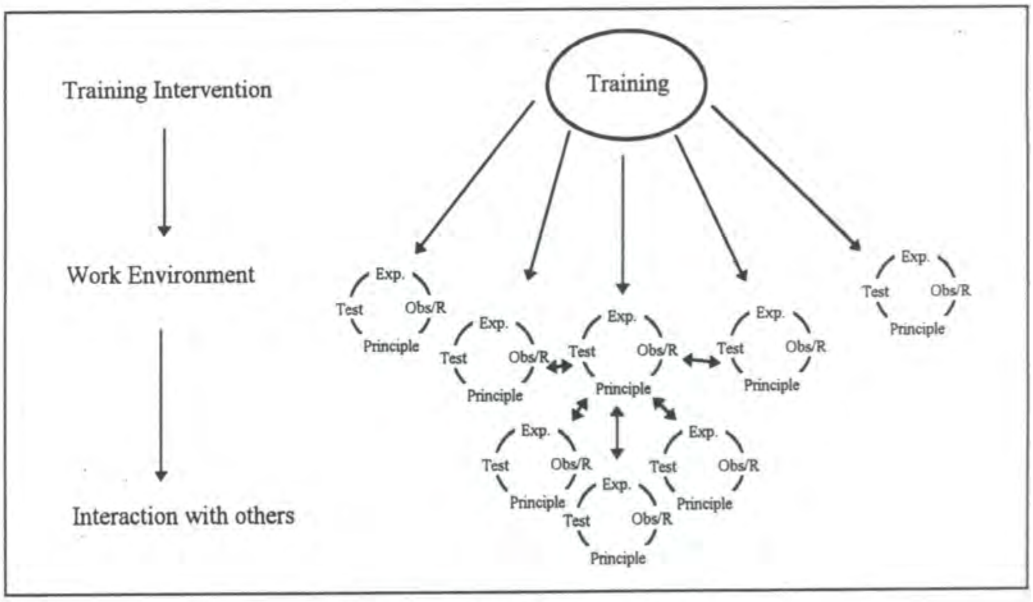

Figure 8-III: Secondary Training Effects

From the empirical study, self evaluation data using behaviourally anchored rating scales was found to be consistent with independent observer evaluation data generated using the same scales. As this has the potential to overcome the cost and practicality and can be integrated into the learning process by providing learners with a reinforcement to the learning (i.e. within a Kolb, 1974, learning cycle type framework), it merits further research and methodology development.

To summarise the conclusions to this study, Kirkpatrick's evaluation framework which has received widespread attention for more than 30 years and become part of the language of training evaluators reinforces a value judgement paradigm focused on outcomes. Training programmes, and in particular those which are long term and intended to bring about organisational change, are far too complex or elaborate so as to be simplified and evaluated in Kirkpatrick's terms. Training is a process concerned with learning and the transfer of skills to an organisational context. The factors or parameters of these processes are dynamic and are known to relate directly to the outcome of training. For evaluation to serve as an improvement tool, evaluation activities must be integrated into the overall learning process, become the tool of the trainer, trainee and manager, and offer relevant real-time information. Evaluation should be conducted within a maximising value paradigm.

8.5 Changing a Paradigm and Organisational Culture

Adoption of a new paradigm creates a high degree of tension between driving and restraining forces within an organisation or a society. For practitioners and researchers alike, this tension is an important consideration. Organisations and societies are described as having cultures which, by way of their nature, are resistant to change, particularly change at the most fundamental level. For the purposes of understanding the magnitude of shifting to a maximising value paradigm, a brief review of organisational psychology literature in this respect is warranted in order to begin to scope the agenda for transition.

8.5.1 Organisational Culture

Originally an anthropological term, culture refers to the underlying values, beliefs and codes of practice that make a community what it is (Fincham and Rhodes, 1998). Cultures reflect the meanings and understandings attributed by members of a society to situations and the solutions that are applied to common problems. Being a member of a society means acquiring core values through the process of growing up and being socialised.

In an organisational sense, Schein (1985) defines culture the 'basic assumptions and beliefs that are shared by members of an organisation'. Bate (1984) argues that a key feature of culture is that it is shared and it refers to ideas, meanings and values people hold in common. Linstead and Grafton (1991) assert that organisational (or corporate) culture is concerned with belonging and conforming and has rites, rituals, stories and values which seduce members and promote collective commitment.

This is assertion is shared by Sathe (1983) who contends that people feel a sense of commitment to an organisation's objectives when they identify with those objectives and experience some emotional attachment to them. The shared beliefs and values that compose culture help generate identification and attachment and equally ideas and values which do not sit within cultural norms are likely to be alienated as they are non-conforming.

Organisational culture is widely conceptualised in the literature, therefore, as an amalgamation of people's attitudes, beliefs and behaviours expressed as 'the way we do things here'. These conceptions can be organised and expressed at three levels (see Figure 8-IV); behaviours; values; and basic assumptions. Behaviours are the visible aspects of the culture, such as the working processes and systems, written and unwritten rules and language and rituals used.

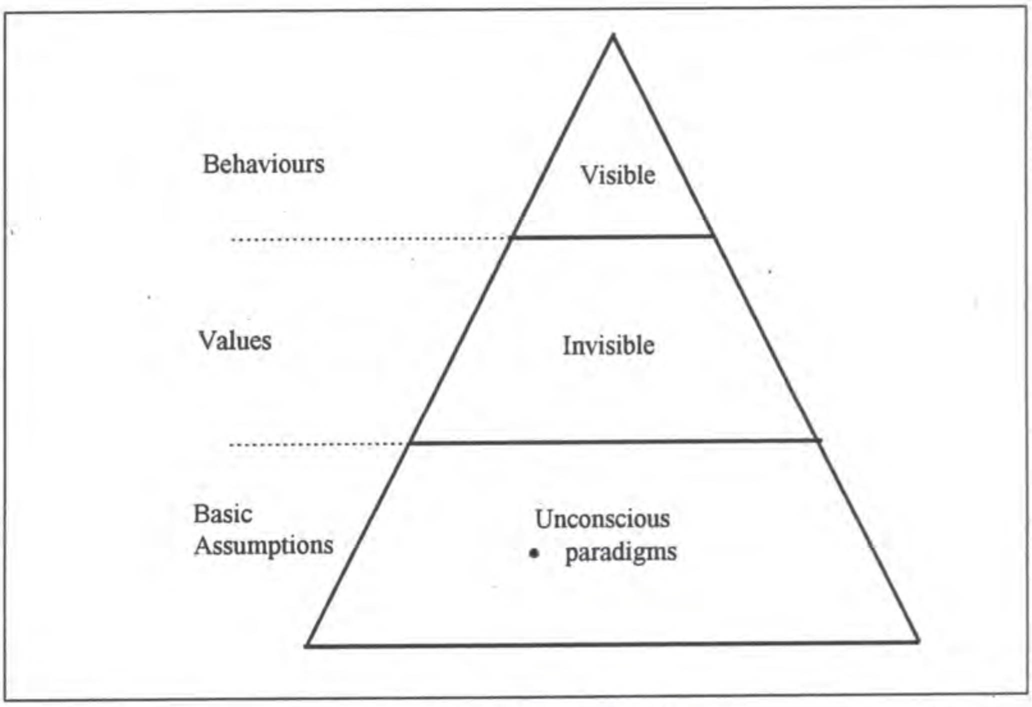

Figure 8-IV: Levels of Organisational Culture

Below the surface of an organisation and beyond what can be seen are the cultural values, beliefs and attitudes, opinions and habits. These are powerful underpining factors of the explicit organisational behaviour. Pascale and Athos (1994) comment that values and beliefs serve to reinforce an organisation as an entity to which individuals identify and belong. Values, beliefs and attitudes are fostered by an organisation and made explicit through behaviour. Conforming behaviour is encouraged, recognised and rewarded which serves to strengthen the underpinning values. The condemnation of non-conforming behaviour prevents the development of counter-values.

Deeper still are the basic assumptions. These are the roots of a culture's history, presuppositions and assumptions which are experienced and passed on by members of the culture as self-evident truths. It is at this level that the concept of paradigms can be applied. Basic assumptions are most difficult to change as they are the filter or mindset through which culture itself is evaluated. Members of an organisation or society are unlikely to recognise the existence of these basic assumptions, or where they do they accept them as absolute truths. Non-members of a society, who are more likely to questions basic assumptions are often repelled by members with suggestions being rejected as out of context.

Fincham and Rhodes (1998) argue that organisational cultures are the natural products of social interaction which are largely unplanned and unpredictable. They evolve and emerge over time and are the residue of countless events and actions.

Whilst the review of organisation culture has been couched mainly in terms of commercial organisations, the same principles can be applied to any society whose members share common beliefs and values. In this sense, proponents of contemporary training evaluation share common beliefs and values, particularly at the basic assumption or unconscious level and can be, and for the purposes of this analysis will be, described as a training evaluation society.

8.5.2 Driving and Restraining Forces of Change

Given this brief review of organisation culture and its influence on individuals within an organisation, and within a training evaluation society, a range of issues emerge for future consideration. To map out the issues involved, force field analysis technique (Lewin, 1951) can be used to serve as an illustration of the driving and restraining forces of an evaluation paradigm shift. Force field analysis is widely used in change management and can be used to help understand most change processes in organisations (Thomas, 1985).

The analysis considered here is by no means comprehensive as it is of insufficient detail. Its purpose is to convey to the reader the broad spectrum of issues to be considered in the context of organisational education and training in its broadest sense, which encompasses the notions of the learning organisation.

In force field analysis, change is characterised as a state of imbalance between driving forces (e.g. new personnel, changing markets, new technology) and restraining forces (e.g. individuals' fear of failure, organisational inertia). To achieve change towards a goal or vision three steps are required; (a) an organisation or society has to unfreeze the driving and restraining forces that hold it in a state of quasi-equilibrium; (b) an imbalance is introduced to the forces to enable the change to take place. This can be achieved by increasing the drivers, reducing the restraints or both; and (c) once the change is complete the forces are brought back into quasi-equilibrium and re-frozen.

Whilst Thomas (1985) notes that force field analysis has been used in various contexts, it is rarely applied to strategy, or in the context of this analysis, philosophy, but argues that it could be used to provide new insights and is therefore a potentially a powerful technique to help an organisation or society realise a strategic or philosophical vision.

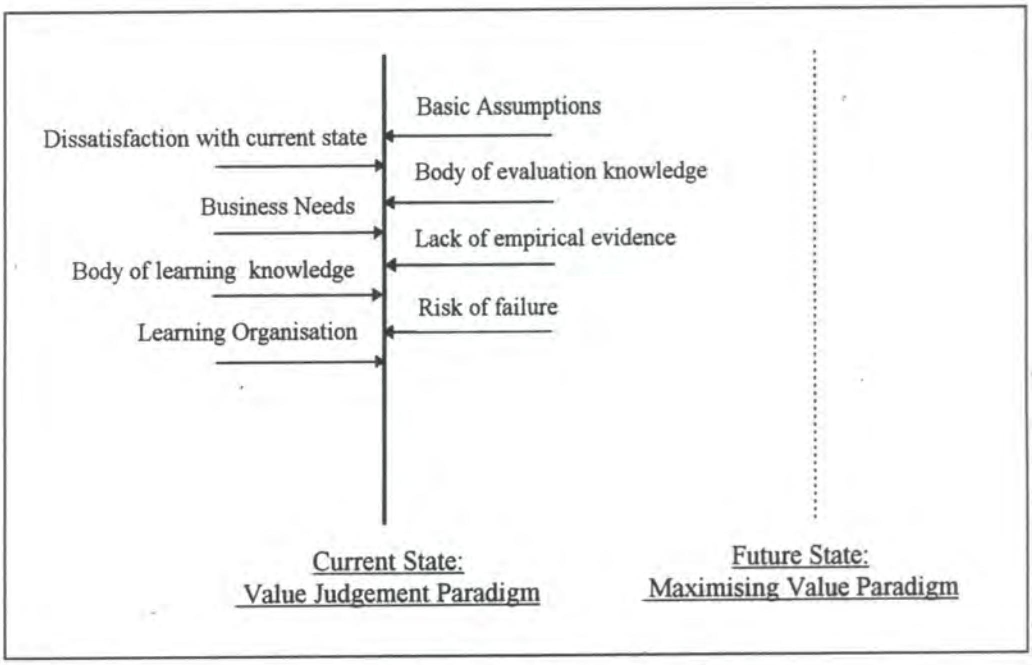

Figure 8-V: Driving and Restraining Forces of Change

Referring to Figure 8-V, the current state of evaluation; value judgement paradigm can be described as being in a state of equilibrium, with respect to a desired future state of a maximising value paradigm. At a macro (general) level this equilibrium can be represented by forces which are acting upon the current state. The analysis is conducted in the context of commercial organisations and against the back-drop of learning, training and training evaluation as conceived in the relevant literature. Essentially it is from the perspective of a training practitioner researcher.

8.5.2.1 Driving Forces

The primary driving forces identified are; (i) dissatisfaction with the current state; (ii) business needs; (iii) body of learning knowledge; and (iv) learning organisation.

The dissatisfaction with the current state is well documented in the literature and has been referenced in chapter 2 of this research. From the researchers own experience of more than 10 years of working in the training function, this dissatisfaction is shared within the training practitioner and commercial business community.

The emphasis on return on training investment (e.g. Brinkerhoff, 1987) is increasing as businesses look to their training departments for evidence of the benefits of training. The importance of learning and training at work (see sections 8.5.3: Importance of Learning at Work for a brief summary) with regard to the performance of commercial operations is serving to drive the training profession into finding more effective and efficient ways of facilitating learning. Evaluation as maximising value facilitates this search.

The body of learning knowledge in the literature is vast and will continue to grow. Better integration with training evaluation is likely to result in synergies which will yield sustained development of our understanding of learning, training and evaluation..

The fourth driving force identified here is the emergence of the learning organisation concept as a positive and popular approach to vocational learning. Similarly to the driving force above, integrating evaluation with this concept as a means to maximise learning is likely to result in synergistic outcomes.

8.5.2.2 Restraining Forces

The primary restraining forces identified are; (i) basic assumptions; (ii) body of evaluation knowledge; (iii) lack of empirical evidence; and (iv) risk of failure.

As previously discussed (sections 8.4.1: Beliefs, Values and Basic Assumptions and 8.5.1: Organisational Culture) the basic assumptions which underpin evaluation as value judgement unite much of the training evaluation society. These basic assumptions securely anchor the society's philosophy and are a powerful restraint of a paradigm shift.

Our current body of knowledge with respect to training evaluation is, by the very nature of its paradigm, implicitly as well as explicit value judgement. Furthermore it is internally integrated with developments based on the work and authority of predecessors. Challenging the existing paradigm cannot be undertaken easily on singular concepts and so it may soon become a challenge of inordinate magnitude.

Within the training community there is no empirical evidence to support this thesis. Without empirical data, persuasive argument with defendents is almost impossible. This relates to the final restraining force identified here, that of risk of failure. Attempting to adopt a new paradigm means there is no guidance from previous research to learn from and few supporters to draw on.

8.5.3 Importance of Learning at Work

Recent training and business literature is punctuated with the pace of change and the need for organisations and individuals to learn more effectively. The importance of learning at work is perhaps best summarised by Mulder, Nijhoff and Brinkerhoff (1995: pp1) who state:

> *"organisations need to (perform effectively) for the sake of continuing their existence in the future. They constantly need to adapt to new circumstances. In the process of focusing on performance, learning plays a critical role."*
>
> *Brackets added*

Mulder et al's statement is as relevant to commercial organisations as it is to any other enterprise which provides a product or service. The relationship between learning and performance is self evident and requires no further discussion, however it is the pace at which organisations and the individuals which make up those organisations learn that is becoming increasingly important.

Organisations who learn quicker and more effectively than their competitors are more likely to be successful and secure long term survival. This is particularly true for the automotive industry where world capacity exceeds demand by over 30%; the world's car-makers have the capacity to build 72 million units annually and yet global consumer demand is less that 50 million units per annum (Autofacts web-site, Sept 1998). It is reasonable to expect that not all of today's car manufacturers will be in existence in 10 or even 5 years time. As the industry consolidates, car manufacture will shift to those areas which have performance advantages, whether those advantages be economic, technological, or political. Only through effective and efficient learning can companies and individuals stay competitive and secure a long term future.

8.6 Final Note on Learning

Apart from serving research purposes with respect to improving how we evaluate training, this project has been a significant learning exercise; to explore the range of knowledge concerned with learning, training, and evaluation. My original draft was in excess of 200,000 words providing descriptive insights into a vast array of theories and practices. The process of summary, which essentially is what this thesis is, a summary of my thoughts, experiences and learning, has taken far more time than I anticipated. The process of summary is a consolidation of one's learning as to describe a concept in a few paragraphs, as opposed to a few pages, requires a very clear understanding of the concept.

Overall I have enjoyed this study and there have been many times when I have experienced a buzz from reading and understanding the work of others, piecing together a string of concepts into coherent knowledge, or getting a chapter finished. However, there have been times when I have felt despair at the size of the undertaking ahead, or disillusioned by the rate of change which occurs in the literature, or frustrated by the imbalance between the amount of time needed and that which I have managed. I will not repeat my statements in my acknowledgement, except to again express my sincere thanks to John McGuiness whom I admire greatly for his wisdom. As for what I will do with my time now - make up for all the time lost with the people I love.

Bibliography

Bibliography

Abercrombie, N., Hill, S., and Turner, B. S. (1984). *Dictionary of Sociology*. Penguin, London.

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behaviour, 3*, 111-149.

Akao, Y. (Ed) (1990). *Quality Function Deployment - Integrating Customer Requirements into Product Design*. Productivity Press, Cambridge, MA, USA.

Aldridge, J. R., and Dale, B. G. (1994). Failure Mode and Effects Analysis. In Dale, B. G. (Ed.) (op.cit.).

Alkin, M. C. (1988). Evaluation - who needs it? Who cares? In Murphy, R. and Torrance, H (Eds). *Evaluating Education: Issues and Methods*. Harper and Row, London.

Alliger, G. M., and Horowitz, H. M. (1989). IBM takes the Guessing out of testing. *Training and Development Journal, 43(4)*, 69-73.

Alliger, G. M. and Janak, E. A. (1987). Kirkpatrick's Levels of Training Criteria: Thirty Years Later. *Personnel Psychology, 42*, 331-342.

American Supplier Institute (1987). *Quality Function Deployment: Implementation Manual for Three Day Workshop*. American Supplier Institute Inc., Dearborn, MI (USA).

Anderson, J. R. (1982). Acquisition of Cognitive Skill. *Psychological Review, 4*, 369-406.

Anderson, J. R. (1983). *The Architecture of Cognition*. Harvard University Press, Cambridge, MA, USA.

Annett, J. and Duncan, K. D. (1967). Task Analysis and Training Design. *Occupational Psychology, 41*, 211-221.

Argyle, M. (1967). *The Psychology of Interpersonal Behaviour*. Penguin, Harmondsworth, Middlesex.

Argyris, C., and Schon, D. A. (1978). *Organisational Learning: A theory of Action Perspective*. Addison-Wesley, Wokingham.

Arnheim, R. (1969). *Visual Thinking*. University of California Press, Berkeley, CA, USA.

Aronson, D. T. & Briggs, L. J. (1983) Contributions of Gagné and Briggs to a Perspective Model of Instruction. In Reigeluth, C. M. (Ed). (op.cit.).

Ashton, D. and Green, F. (1996). *Education, Training and the Global Economy*. Edward Elgar Publishing, Cheltenham.

ASTD (1990). Evaluation Framework, Design, and Reports: *Supplement to the Training and Development Journal (July '90)*, American Society for Training and Development, USA.

Atkinson, R. L., Atkinson, R. C., Smith, E. E., and Bem, D. J. (1993). *Introduction to Psychology*. Harcourt Brace Jovanovich, Orlando, USA..

Ausubel, D. P. (1968). *Educational Psychology; A Cognitive View*. Holt, Rinehart and Winston, New York, USA.

Ayto, J. (1991). *Dictionary of Word Origins*. Bloomsbury Publishing, London.

Bandura, A. (1977). *Social Learning Theory*. Prentice-Hall, Englewood Cliffs, NJ, USA.

Bandura, A., Adams, N. E., & Beyer, J. (1977). Cognitive processes mediating behavioural change. *Journal of Personality and Social Psychology*, 35, pp 125-139.

Barham, K., Fraser, J., and Heath, L. (1988). *Management for the Future*. Ashridge Management Research Group, Berkhamstead.

Barker, J. A. (1987). *The Business of Paradigms: Discovering the Future*, Charthouse International, London.

Basarab, D. J. (1994). Evaluating the Creative Manager Program. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Basarab, D. J. (1994). Foreword. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Basarab, D. J. (1998). Using Quality to Drive Evaluation: A Case Study. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Basarab, D. J., & Root, D.K. (1994). *The Training Evaluation Process*. Kluwer Academic Publishers, London.

Bate, P. (1984). The Impact of Organizational Culture on Approaches to Organizational Problem Solving, *Organization Studies*, 5(1), 43-66.

Belbin, M. (1993). *Team Roles at Work*, Butterworth-Heinemann, Oxford.

Bergan, J. R. (1980). The Structural Analysis of Behavior: An Alternative to the Learning Hierarchy Model. *Review of Educational Research*, 50, 625-646.

Berlyne, D. E. (1965). Motivational Problems raised by exploratory and epistemic behaviour. In Koch, S. (Ed.) *Psychology: A Study of a Science (Vol. 5)*. McGraw-Hill, New York, USA.

Bern, A. T. (1988). *Team Oriented Problem Solving: Pilot Training Course Report*. In-Company correspondence.

Bingham, W. V. D., and Moore, B. V. (1959). *How to Interview*. Harper and Row, New York, USA.

Bloom, B. S. (ed.) (1956).*Taxonomy of Educational Objectives: Handbook I: The Cognitive Domain*. Longman, London.

Bond, S. (1994). Evaluating a Training Program on Evaluation Skills. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Bossert, J.L. (1991). *Quality Function Deployment: A Practitioner's Approach*. ASQC Press, Milwaukee, WI, USA.

Bower, G. H. and Hilgard, E. R. (1981). *Theories of Learning*. Prentice-Hall, Englewood Cliffs, NJ, USA.

Bramley, P. (1991), *Evaluating Training Effectiveness: Translating Theory into Practice*. McGraw-Hill, Maidenhead, Berks.

Bramley, P. (1997). *Evaluating Training*. Institute of Personnel and Development, London.

Brandenberg, D. C. and Smith, M. E. (1991). *Evaluation of Corporate Training Programs ERIC / TME Report*, ERIC Clearing House on Tests, Measurement, and Evaluation, Princeton, New Jersey, USA.

Brandes, D. V. (1985). *An Illuminative Evaluation of an Alcohol Education Project*. PhD Thesis, University of Durham.

Braithwaite, R. (1955). *Scientific Explanation*. Cambridge University Press, Cambridge.

Breakwell, G. M., Foot, H., & Gilmour, R. (Eds.) (1988). *Doing Social Psychology: Laboratory and Field Exercises*. Press Syndicate of the University of Cambridge, Cambridge.

Briggs, L. J. (1970). *Handbook of Procedures for the design of Instruction*. American Institute for Research, Pittsburgh, USA.

Briggs, L. J. (1977) *Instructional Design: Principles and Applications*. Educational Technology Publications, Englewood Cliffs, NJ, USA.

Briggs, L. J. and Wager, W. W. (1981). *Handbook of Procedures for the design of Instruction*. Educational Technology Publications, Englewood Cliffs, NJ, USA.

Briggs-Myers, L. and Myers, P. B. (1993). *Gifts Differing*. Consulting Psychological Press, Pao Alto, CA (USA).

Brinkerhoff, R. O. (1991). *Achieving Results from Training*. Jossey-Bass, Oxford.

Brinkerhoff, R. O. (1995). Evaluation Models and Strategies: Introduction. In Mulder, M., Nijhof, W. J., and Brinkerhoff, R. O. (Eds.) *Corporate Training for Effective Performance*. Kluwer Academic Publishers, London.

Brittle, R. J. (1991-1996). Unpublished research log.

Broad, M. L. and Newstrom, J. W. (1992). *Transfer of Training*. Addison-Wesley, Reading, MA, USA.

Brookfield, S. D. (1986). *Understanding and Facilitating Adult Learning*. Open University Press, Milton Keynes.

Brown, J. S., Collins, A., and Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, 18, 32-42.

Brown, L. (Ed.) (1993). *The New Shorter Oxford English Dictionary on Historical Principles (Vols 1 & 2)* Oxford University: Clarendon Press, Oxford.

Brown, S. M. (1998). The Changing Context of Practice. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Brown, S. M., and Seidner, C. J. (Eds.) (1998). *Evaluating Corporate Training: Models and Issues*. Kluwer Academic Publishers, London.

Bruner, J. S. (1960). *The Process of Education*. Vintage Books, New York, USA.

Bruner, J. S. (1966). *Towards a Theory of Instruction*. The Belknap Press of Harvard University Press, Cambridge, MA, USA.

Buckley, R. and Caple, J. (1992). *The Theory and Practice of Training*. Gower, Aldershot, Hants.

Burgess, R. G. (Ed.) (1989). *The Ethics of Educational Research*. Falmer Press, Lewes.

Burgoyne, J. G. (1992). *Stakeholder Analysis in Qualitative Methodology*. Sage, London.

Burgoyne, J. G. (1994) Stakeholder Analysis. In Cassell, C. and Symon, G. (Eds.) *Qualitative Methods in Organisational Research*. Sage, London.

Buzan, T, and Buzan, B. (1993). *The Mind Map Book*. BBC Books, London.

Cairns, L. (1989). Reinforcement. In Hargie, O. (Ed.) (op.cit.).

Calvert, G., Mobley, S., and Marshall, L. (1994). Grasping the Learning Organization. *ASTD Training and Development Journal,* June, 39-43.

Camp, R. R., Blanchard, P. N., and Huszco, G. E. (1986). *Toward a More Organizationally Effective Training Strategy and Practice*. Prentice-Hall, Englewood Cliffs, NJ, USA.

Campbell, J. P. (1971). Personnel Training and Development. *Annual Review of Psychology,* 22, 565-602.

Campbell, J. P. (1989). The Agenda for Theory and Research. In Goldstein, I. L. & Associates (op.cit.).

Cannon-Bowers, J. A., and Tannenbaum, S. I. (1991). Toward an Integration of Training Theory and Technique. *Human Factors, 33(3)*, 281-292.

Carnevalle, A. P. & Schulz, E. R. (1990). Evaluation Framework, Design, and Reports. Supplement to the *Training and Development Journal,* July.

Cascio, W.F. (1987). *Applied Psychology in Personnel Management*. Englewood Cliffs, NJ: Prentice-Hall.

Cavan, S. (1977). *The American Journal of Sociology*, 83, 110. Boston, MA, USA.

Checkland, P. B. (1981). *Systems Thinking, Systems Practice*. John Wiley and Sons Ltd, London.

Clarke, J. P. (1994). Evaluating a Training Course on Performance Appraisal and Training. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Clement R. W. (1981). Testing the Hierarchy Theory of Training Evaluation: An Expanded Role for Trainee Reactions, *Public Personnel Management Journal*.

Cohen, L., and Manion, L. (1994). *Research Methods in Education*. Routledge, London.

Cohen, M. R., and Nagel, E. (1934). *An Introduction to Logic and the Scientific Method.* Harcourt Brace Jovanovich, New York, USA.

Collins, A. and Stevens, A. L. (1983). A Cognitive Theory of Inquiry Teaching. In Reigeluth, C.M. (Ed.) (op.cit.).

Connor, R. (1985). International and Domestic Evaluation: Comparisons and Insights. In Patton, M. Q. (Ed.) (1985). *Culture and Evaluation*. Jossey Bass, San Francisco, CA, USA.

Cormier, S. M., and Hagman, J. D. (1987). *Transfer of Learning: Contemporary Research and Applications*. Academic Press, London.

Crannell, C.W. (1956). Transfer in problem solution as related to the type of training. *The Journal of General Psychology*, 54, 3-14.

Cronbach, L. J. (1963). Course Improvement through evaluation. Reprinted in Madaus, G. F., Scriven, M. S., and Stufflebeam, D. L. (Eds.) (1996) (op.cit.).

Crosby, P. B. (1979). *Quality is Free: The Art of Making Quality Certain*. McGraw-Hill, New York, NY, USA.

Crosby, P. B. (1984). *Quality without Tears: The Art of Hassle-Free management*. McGraw-Hill, New York, NY, USA.

Cummings, O. W. (1998). What Stakeholders Want to Know. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Dahl, H. L. (1987). Return on Investment. In Craig, R. L. (Ed.) (1987). *Training and Development Handbook: A Guide to Human Resource Development*. McGraw-Hill, New York, USA.

Dale, B. G., Shaw, P. (1989). *Failure Mode and Effects Analysis: A study of its use in the Motor Industry*. Occasional Paper 8904. Manchester School of Management. University of Manchester.

Dale, B.G. (Ed.) (1989). *Tools and Techniques for TQM: Proceedings of the First Conference*. IFS Publications, Kempston, Bedford.

Dale, B.G. (Ed.) (1994). *Managing Quality*. Prentice Hall, Hemel Hempstead, Herts.

Dalziel, S. (1991). Organizational training Needs. In Prior, J. (Ed.) 1991). *Handbook of Training and Development*. Gower, Aldershot, Hants.

Dane, F. C. (1990). *Research Methods*. Brooks/Cole Publishing Company, Belmont, CA, USA.

De Bono, E. (1991) . *Teaching Thinking*. Penguin, London.

Dembo, M. H. (1981). *Teaching for Learning*. Goodyear Press, London.

Deming, W. E. (1975), The Logic of Evaluation. In Guttentag, M. and Streuning, E. L. (Eds.) (1975). *Handbook of Evaluation Research (vol. 1)*. Sage Publishing, Beverly Hills, CA, USA.

Deming, W. E. (1982). *Out of the Crisis*. Massachusetts Institute of Technology, Cambridge, MA, USA.

Deming, W. E. (1983). *Quality, Productivity, and Competitive Position*, Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, MA, USA.

Deming, W. E. (1993). *The New Economics: for Industry, Government, Education*, Massachusetts Institute of Technology, Cambridge, MA, USA.

Department of Employment Training Agency (1989*), Training in Britain: A Study of Funding, Activity and Attitudes*. HMSO, London.

Dewey, J. (1933). *How We Think*. Heath, Boston, MA, USA.

Douglas, J. D. (1976). *Investigative Social Research: Individual and Team Field Research*, Sage, Beverly Hills, CA, USA.

Douglass, B. and Moustakas, C. (1984). *Heuristic Inquiry: The Internal Search to Know*. Centre for Humanistic Studies, Detroit, MI, USA.

Dubin, S., and Okun, M. (1973). Implications of Learning Theories for Adult Education, *Adult Education* (USA), 24(1), 3-19

Easterby-Smith, M. (1994). *Evaluating Management Development, Training and Education*. Gower Publishing, Aldershot, Hants.

Eisner, E. W. (1976). Educational Connoisseurship and Criticism: Their Form and Functions in Educational Evaluation. *Journal of Aesthetic Education*, 10, 145-50.

Enever, L. and Harlen, W. (1972). *With Objectives in Mind*. MacDonald, London.

Feather, N. T. (1975). *Values in Education and Society*. The Free Press, New York, USA.

Feigenbaum, A. V. (1961). *Total Quality Control: Engineering and Management*. McGraw-Hill, London.

Feigenbaum, A. V. (1986). *Total Quality Control*. McGraw-Hill, New York, NY, USA.

Feigenbaum, E. A., & Feldman, J. (1963). *Computers and Thought*. McGraw-Hill, New York, USA.

Filmer, P. (1973). *New Directions in Sociological Theory*. MIT Press, Cambridge, MA, USA.

Filstead, W. J. (1979). Qualitative Methods: A Needed Perspective in Evaluation Research. In Cook, T. D. and Reichardt, C. S. (Eds.*). Qualitative and Quantitative Methods in Evaluation Research*. Sage, Beverly Hills, USA.

Fincham, R. and Rhodes, P. S. (1998). *The Individual, Work and Organization* (2^{nd} ed.). Oxford University Press, New York, NY, USA.

Fitts, P. M. (1962). Factors in Complex Skill Training. In Glaser, R. (Ed*.) Training Research and Education*. University of Pittsburg. (Reprinted in 1965, Wiley, New York, USA.)

Fitz-enz, J. (1993). *Benchmarking Staff Performance*. Jossey-Bass, San Francisco, CA.

Flanagan (1954). The Critical Incident technique. *Psychological Bulletin, 51*.

Fleishman, E. A., and Mumford, M. D. (1989). Individual Attributes and Training Performance. In Goldstein, I. L. and Associates (1989) (op.cit.).

FMC (1984). Worldwide Failure Mode and Effects Analysis Guidelines. Ford Motor Company, Dearborn, MI, USA.

FMC (1987). Internal document (Sept, 22): *Working Group Reports - Inhibitors, Critical Success Factors and Opportunity Maps*. Quality Strategy Office, Ford of Europe, Brentwood, Essex.

FMC (1990a). Internal document: *Engineering Needs Analysis Report* (Sept). Company Motor Company Ltd, Brentwood, Essex.

FMC (1990b). *World-wide Quality System Standard Q-101*. Corporate Quality Office, Ford Motor Company Inc., Dearborn, MI (USA).

FMC (1991a). Internal memo (April 9): *Training Case Study - Specification*. Ford Motor Company Ltd, Brentwood, Essex.

FMC (1991b). Internal memo (Sept 12): *Programme Development Team Membership, Roles and Responsibilities*. Ford Motor Company Ltd, Brentwood, Essex.

FMC (1993). *Evaluation Strategy: North American Operations*. In-company correspondence. Company Motor Company, Dearborn, MI, USA.

FMC (1995). Internal document: *Programme Budget Analysis Review*. ET&D, Ford Motor Company Ltd, Brentwood, Essex.

FMC (1997). Internal memo (Nov 20): *Considerate Use of the Company E-mail System*. Quality and Process Leadership, Ford Motor Company Ltd, Brentwood, Essex.

FMC and The Sporting Body Mind (1989). *Team Oriented Problem Solving*. Newgate Press, London.

Forgas, J. P. (1982). Episode Recognition: Internal Representation of Interaction Routines. In Berkowitz, L. (Ed.). *Advances in Experimental Social Psychology*. Academic Press, Newe York.

Freitag, E. (1994). Implementing the Kirkpatrick Model as an Up-front Analysis and Evaluation Tool. In Kirkpatrick, D. L. (Ed.) (op.cit.).

French, H. P. (1989). *An assessment of the pre-registration preparation of nurses as an educational experience* [pp39]. PhD Thesis. School of Education, The University of Durham.

French, S. H. (1953). Measuring Progress to Industrial Relations Objectives. *Personnel, 30,* 338-347.

Gagné, R. M., Briggs, L. J. & Wager, W. W. (1993). *Principles of Instructional Design*. Harcourt Brace Jovanovich, Orlando, FL, USA.

Gagné, R. M. (1970). *The Conditions of Learning*. Holt, Rinehart & Winston, New York, USA.

Gagné, R. M. (1977). *The Conditions of Learning*. Holt, Rinehart and Winston, New York, USA.

Gagné, R. M. and Briggs, L. J. (1979). *Principles of Instructional Design*. Holt, Rinehart and Winston, New York, USA.

Gardin, G. R., & Sitterly, T.E. (1972). *Degradation of Learned Skills - A review and annotated bibliography*. Report Number D180-15080-1. Boeing Company. In Patrick, J. (1992) (op.cit.).

Gick, M. L., and Holyoak, K. J. (1983). Schema Induction and Analogical Transfer. *Cognitive Psychology*, 15, 1-38.

Glass, G. V. and Ellett, F. S. (1980). Evaluation Research. *Annual Review Psychology*, 211-228, Annual Reviews Inc., CA, USA.

Goldbeck, R.A., Bernstein, B.B., Hillix, W.A., & Marx, M.H. (1957). Application of the half split technique to problem solving tasks. *Journal of Experimental Psychology*, 53, 330-338.

Goldstein, I. L. & Associates (1989). *Training and Development in Organizations*. Jossey-Bass, San Francisco, CA, USA.

Goldstein, I. L. (1986). *Training in Organisations: Needs Assessments, Development, and Evaluation*. Brooks/Cole, Monterey, CA, USA.

Goldstein, I. L. (1993). *Training in Organisations*. Brooks/Cole, Monterey, CA, USA.

Golembiewski, R. T., Billingsley, K., and Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioural Science, 12*, 133-157.

Gordon, J. (1997). The HRD Hall of Fame, *Training Magazine*; February.

Gould, S. J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History*. Penguin, London.

Gropper, G. L. (1974). *Instructional Strategies*. Educational Technology Publications, Englewood Cliffs, NJ, USA.

Gropper, G. L. (1975). *Diagnosis and Revision in the development of Instructional Materials*. Educational Technology Publications, Englewood Cliffs, NJ, USA.

Gropper, G. L. (1983). A Behavioural Approach to Instructional Prescription. In Reigeluth, C.M. (Ed.) (op.cit.).

Gross, R. D. (1992). *Psychology: Science of Mind and Behaviour*. Hodder and Stoughton, London.

Grossberg, S. (1978) A Theory of Human Memory: Self-organization and Performance of Sensory-motor codes, Maps and Plans. In R. Rosen & F. Snell (Eds.). *Progress in Theoretical Biology (Vol. 5)*. Academic Press, New York, USA.

Grove, D. M., and Davies, T. P. (1992). *Engineering Quality and Experimental Design*. Longman, Harlow, Essex.

Guba, E. G. & Lincoln, Y. S. (1981). *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. Jossey-Bass, San Francisco, CA, USA.

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Sage Publishing, London.

Guttentag, M. and Streuning, E. L. (Eds.) (1975). *Handbook of Evaluation Research (Vol. 1)*. Sage Publishing, Beverly Hill, CA, USA.

Hall, D. T., and Goodall, J. G. (1986). *Human Resource Management: Strategy, Design, and Implementation*. Scott, Foresman and Company, Glenview, Il, USA.

Hamblin, A. C. (1974). *Evaluation and Control of Training*. London: McGraw-Hill.

Hammersley M. (Ed.) (1993). *Social Research; Philosophy, Politics and Practice*. Sage, London.

Hammond, V., & Wille, E. (1991). The Learning Organization. In Prior, J. (Ed.). *Handbook of Training and Development*. Gower, Aldershot, Hants.

Hargie, O. (Ed.) (1988). *A Handbook of Communication Skills*. Routledge, London.

Harman, H. (1976). *Modern Factor Analysis*. Chicago University Press, Chicago, Il, USA.

Harper, W. M. (1983). *Statistics*. Macdonald and Evans, Plymouth.

Harris, D., Bell, C. (1994). *Evaluating and Assessing for Learning*. Kogan Page, London.

Hawthorne, E. M. (1987). *Evaluating Employee Training Programs*. Quorum Books, Westport, CT (USA).

Henerson, M. E., Morris, L. L., and Fitz-Gibbon, C. T. (1987). *How to Measure Attitudes*. Sage Publications, Newbury Park, CA, USA.

Henshall, E. J. (1989). Ford - SPC is Only Part of the Picture. *Quality Today Journal*, October.

Henshall, E. J. (1990). The Application of Quality Techniques: Achieving a Balance. *Total Quality Magazine*, November.

Herzberg, F., Mausner, B., and Snyderman, B. B. (1959). *The Motivation to Work*. John Wiley and Sons, New York, NY, USA.

Hill, S. (1994). Quality Circles in Britain. In Clark, H., Chandler, J. and Barry, J. (Eds). *Organisation and Identities: Text and Readings in Organisational Behaviour*. Chapman and Hall, London.

Hinrichs, J. R. (1976). Personnel Training. In Dunnette, M. D. (Ed.). *Handbook of Industrial and Organizational Psychology*. Rand McNally, Stokie, Il, USA.

Holcombe, J. (1994). Evaluating a Training Programme for District Managers. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities*. Addison-Wesley, Reading, MA, USA.

Holton, A. (1996). The Flawed Four-level Evaluation Model. *Human Resource Development Quarterly*, 7(1); 23-26.

Homa, D., and Cultice, J. (1984). Role of Feedback, Category Size, and Stimulus Distortion on the Acquisition and Utilization of Ill-defined Categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 83-94.

Honey, P. and Mumford, A. (1986). *Using Your Learning Styles*. Peter Honey, Ardingly House, 10 Linden Avenue, Maidenhead, Berks.

Honey, P. and Mumford, A. (1992). *A Manual of Learning Styles*. Peter Honey Publications, Maidenhead, Berks.

Honey, P. and Mumford, A. (1996). *How to Manage your Learning Environment: Making Learning a priority at work*. Peter Honey Publications, Maidenhead, Berks.

Honey, P. (1997). *Annual Calendar: Learning*. Peter Honey Publications, Maidenhead, Berks.

House, E. R. (1978). Assumptions Underlying Evaluation Models. *Educational Researcher*, March, 4-12. American Educational Research Association, Washington, D.C. USA.

House, E. R. (Ed.) (1986). *New Directions in Educational Evaluation*: Falmer Press, Philadelphia, PA, USA.

Hunt, E. (1989). Cognitive Science: Definition, Status, and Questions. *Annual Review of Psychology*, 40, 603-629.

Ishikawa, K. (1982). *Guide to Quality Control.* Quality Resources, White Plains, NY, USA.

Ishikawa, K. (1985). *What is Total Quality Control?: The Japanese Way.* Prentice Hall, Englewood Cliffs, NJ, USA.

Jackson, T. E. (1989). *Evaluation: Relating Training to Business Performance*. Kogan Page, London.

Jackson, T. E. (1991). *Measuring Management Performance: A developmental approach for trainers and consultants*. Kogan Page, London.

Jaguar Cars (1986). *FMEA: A Guide to Analysis*. Coventry.

Johnson, K.G. (1997). *A Study into the Use of Process Failure Mode and Effects Analysis (PFMEA) and its Effectiveness as a Problem Prevention Technique*. M.Sc. Dissertation; University of Bradford.

Joint Committee on Standards for Educational Evaluation (1981). Standards for the Evaluation of Educational Programs, Projects and Materials. In Guba, E. G. & Lincoln, Y. S. (1989) (op.cit.).

Jones, A.M., and Hendry, C. (1994) The Learning Organization: Adult Learning and Organisational Transformation. *British Journal of Management*, Vol. 5, 153-162.

Joseph, S. A., Brewin, C. R., Yule, W., and Williams, R. (1993). Causal Attributions and Post-traumatic Stress in Adolescents. *Journal of Child Psychology and Psychiatry, 34*, 247-253.

Juran, J.M. (1989). *Juran on Leadership for Quality: An Executive Handbook.* Macmillan Free Press, New York, USA.

Juran, J.M. (Ed.) (1988). *Quality Control handbook.* McGraw-Hill, New York, USA.

Juran, J.M. (1992). *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services*. Macmillan Free Press, New York, NY, USA.

Kandola, P. (1993). *Becoming a Learning Organisation: A Toolkit*. Institute of Training and Development, Marlowe.

Kano, N., and Takahashi, F. (1979). *The Motivator - Hygiene Factor in Quality*. Japanese Society for Quality Control, Abstracts of the 9th Annual Presentation Meeting.

Kaposi, A. and Myers, M. (1994). *Systems, Models and Measures*. Spronger-Verlag, London.

Kearsley, G. (1982). *Costs, Benefits, and Productivity in Training*. Addison-Wesley, Reading, MA, USA.

Keller, J. M. (1983). *Motivational Design of Instruction*. In Reigeluth, C.M. (Ed.) (op.cit.).

Kelly, G. (1955). *The Psychology of Personal Constructs*. Norton, New York, USA.

Kepner, C. H. and Tregoe, B. B. (1981). *The New Rational Manager*. Princeton Research Press, Princeton, NJ, USA.

Kerlinger, F. N. (1986). *Foundations of Behavioural Research*. Holt, Rinehart and Winston, New York, USA.

Kidder, P. J. and Rouiller, J. Z. (1997). Evaluating the Success of a Large Scale Training Effort. *National Productivity Review*, Spring, 79-89.

Kim, D.H. (1993). The Link Between Individual and Organisational Learning. *Sloan Management Review*, Fall, 37-50. Massachusetts Institute of Technology, MA, USA.

Kirkpatrick, D. L. (1959a). Techniques for Evaluating Training Programs. *Journal of America Society for Training and Development, 13* (12), 3-9.

Kirkpatrick, D. L. (1959b). Techniques for Evaluating Training Programs: Part 2 - Learning. *Journal of America Society for Training and Development, 13* (12), 21-26.

Kirkpatrick, D. L. (1960a). Techniques for Evaluating Training Programs: Part 3 - Behavior. *Journal of America Society for Training and Development, 14* (1), 13-18.

Kirkpatrick, D. L. (1960b). Techniques for Evaluating Training Programs: Part 4 - Results. *Journal of America Society for Training and Development, 14* (2), 28-32.

Kirkpatrick, D. L. (Ed.) (1975). *Evaluating Training Programs*. American Society for Training and Development, Madison, WI, USA.

Kirkpatrick, D. L. (1994). Evaluating a Training Program on Developing Supervisory Skills. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Kirkpatrick, D. L. (Ed.) (1994). *Evaluating Training Programs: The Four Levels*. Berrett-Koehler, San Francisco.

Kirkpatrick, D. L. (1996). Revisiting Kirkpatrick's Four-Level Model. *Training and Development Journal*, 50(1), 54-60.

Kirkpatrick, D. L. (1998). The Four Levels of Evaluation. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Knowles, M. (1990). *The Adult Learner: A Neglected Species*. Gulf Publishing, London.

Koestler, A. (1964). *The Art of Creation*. Hutchinson, London.

Koffka, K. (1924). *The Growth of the Mind*. Translated by R. M. Ogden. Kegan Paul, Trench, Tubner, London.

Kolakowski, L. (1993). An Overall View of Positivism. In Hammersley, M. (Ed.) (op.cit.).

Kolb, D. (1984). *Experiential Learning*. Prentice-Hall, Englewood Cliffs, CA, USA.

Krathwohl, D. R. (1985). *Social and Behavioural Science Research*. Jossey-Bass, Oxford.

Krathwohl, N. (Ed.) (1964).*Taxonomy of Educational Objectives: Handbook II - the Affective Domain*. Longman, Harlow, Essex.

Krijger, N. L., and Pol, S. M. (1995). Changing Leadership Style: a Training Model for Lasting Effects. In Mulder, M., Nijhof, W. J., and Brinkerhoff, R. O. (Eds.) *Corporate Training for Effective Performance*. Kluwer Academic Press, London.

Krippendorff, K. (1981). *Content Analysis: An Introduction to its Methodology*. Sage, London.

Krueger, R. A. (1988). *Focus Groups*. Sage, London.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, Il, USA.

Kyberg, H. E. (1984). *Measurement Theory*. Cambridge University Press, Cambridge.

Landa, L. (1974). *Algorithmization in Learning and Instruction*. Educational Technology Publications, Englewood Cliffs, CA, USA.

Landa, L. (1983). The Anglo-heuristic theory of Instruction. In Reigeluth, C.M. (Ed.) (op. cit.).

Larson, C. E. and Lafasto, F. M. J. (1989). *Teamwork; What must go right/ What can go wrong*. Sage Publications, Newbury Park, CA, USA.

Latham, G. P. (1988). Human Resource Training and Development. *Annual Review of Psychology, 39*, 545-582.

Lawler, P. A. (1998). The Ethics of Evaluating Training. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Lewin K. (1951). *Field Theory in Social Science*, Harper and Row, New York, NY, USA.

Linstead, S. L. and Grafton Small, R. G. (1991). On Reading Organizational Culture. *Organizational Studies*, 13(3).

Lohman, D. F., and Snow, R. E. (1984). Toward a Theory of Cognitive Aptitude for Learning from Instruction. *Journal of Educational Psychology,* 76, 374-376.

MacKay, D. G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behaviour. *Psychological Review*, 89(5), 483 - 506.

MacRae, A. W. (1994). *Models and Methods for the Behavioural Sciences*. British Psychological Society, Leicester.

Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (1996). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Kluwer-Nijhoff Publishing, Boston, MA, USA.

Mager, R. F., and Pipe, P. (1990). *Analysing Performance Problems*. Kogan Page, London.

Maier, N.R.F. (1973). *Psychology in Industrial Organizations*. Houghton Mifflin, Boston, MA., USA.

Marshall, R. and Tucker, M. (1992). *Thinking for a living*. Basic Books, New York, USA.

Maslow, A. H. (1954). *Motivation and Personality*. Harper and Row, New York, USA.

Mayo, E. (1946). *Human Problems of an Industrial Civilisation*. Macmillan, London.

McClelland, D. C. (1976). *The Achieving Society*. Irvington Publications, New York, USA.

Merrill, M.D. (1983). Component Display Theory. In Reigeluth, C. M. (Ed.) (op.cit.).

Merton, R., Riske, M., and Kendall, P. L. (1956). *The Focused Interview*, Free Press, New York, USA.

Micklethwaite, J., and Wooldridge, A. (1997). *The Witch Doctors*. Mandarin, London.

Moore, C. A., and Sedner, C. J. (1998). Organizational Strategy and Training Evaluation. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Morrison, K. (1993). *Planning and accomplishing school-centred evaluation*. Peter Francis Publishers, Dereham, Norfolk.

Mosser, C. A., and Kalton, G. (1971). *Survey Methods in Social Investigation*. Dartmouth Publishing, Aldershot, Hants.

Munro-Faure, L., and Munro-Faure, M. (1992). *Implementing Total Quality Management*. Pitman Publishing, London.

Neale, F. (1991). *The Handbook of Performance Management*. Institute of Personnel Management, London.

Nemoto, M. (1987). Total Quality Control for management: Strategies and Techniques from Toyota and Toyoda Gosei (Translated by and edited by Lu, D.). Prentice Hall, Englewood Cliffs, NJ, USA.

Newby, A. C. (1992). *Training Evaluation Handbook*. Gower, Aldershot, Hants.

Noe, R. A. (1986). Trainees' Attributes and Attitudes: Neglected Influences on Training Effectiveness. *Academy of Management review*, 11(4), 736-749.

Noe, R. A., & Schmitt, N. (1986). The Influence of Trainee Attitudes on Training Effectiveness: Test of a Model. *Personnel Psychology*, 39, 497-523.

Oakland, J. S. (1986). *Statistical Process Control: A practical Guide*. Heinemann, London.

Oakland, J. S. (1993). *Total Quality Management: The Route to Improving Performance*. Butterworth-Heinemann, Oxford.

O'Hara, P. (1994). Evaluating a Training Program for Nonexempt Employees. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Oppenheim, A. N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. Pinter Publishers Ltd, London.

Ornstein, R. (1972). *The Psychology of Consciousness*. Freeman and Co. San Francisco, USA.

Ostroff, C, and Ford, J. K. (1989). Assessing Training Needs: Critical Levels of Analysis. In Goldstein, I. L. and Associates (op.cit.).

Paquet, B, Kasl, E., Weinstein, L., and Waite, W. (1994) The Bottom Line. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Pareto, V. (1972). *Manual of Political Economy*. Macmillan, London.

Parlett, M. R. and Dearden, G. J. (Eds.) (1977). *Introduction to Illuminative Evaluation: studies in higher education*. Pacific Surroundings Press, Cardiff-by-the-Sea, CA, USA.

Parlett, M. R. and Hamilton, D. F. (1972). *Evaluation as Illumination: A New Approach to the Study of Innovatory Programs*. University of Edinburgh, Centre for Research in the Educational Sciences, Occasional Paper No. 9.

Parlett, M. R. in Reason, P. and Rowan, J. (Eds.) (1981). *Human Inquiry: A Sourcebook of New Paradigm Research*. Chichester: John Wiley & Sons.

Pascale, R. and Athos, A. (1994). Corporate Cultures. In Clark, H., Chandler, J. and Barry, J. (Eds). *Organisation and Identities: Text and Readings in Organisational Behaviour*. Chapman and Hall, London.

Patrick, J. (1992). *Training: Research and Practice*. Academic Press, London.

Patton, M. Q. (1978). *Utilization-Focused Evaluation*. Sage Publications Ltd, London.

Patton, M. Q. (1982) *Practical Evaluation*. Sage Publications Ltd, London.

Patton, M. Q. (1990), *Qualitative Evaluation and Research Methods*. Sage Publications, London.

Pavlov, I. P. (1927). *Conditioned Reflexes*. Clarendon Press, London.

Pearson, E. S., and Hartley, H. O. (Eds.) (1954). *Biometrika Tables for Statisticians (Vol. 1)*, Cambridge University Press, Cambridge.

Peddler, M., Burgoyne, J., and Boydell, T. (1991). *The Learning Company*. McGraw-Hill, Maidenhead.

Peters, T. (1988). *Thriving on Chaos: Handbook for a Management Revolution*. MacMillan, London.

Peters, T., and Waterman, R. H. Jr. (1982). *In Search of Excellence*. Harper and Row, New York, NY, USA.

Petersen, D. E., and Hillkirk, J. (1991). *Teamwork: New Management Ideas for the Nineties*. Gollancz, London.

Phillips, J. J. (1991). *Handbook of Training Evaluation and Measurement Methods*. Gulf Publishing Co., Huston, Texas, USA.

Phillips, J. J. (1994). *In Action: Measuring Return on Investment*. American Society for Training and Development, Alexandria, Virginia, USA.

Phillips, J. J. (1998). Level Four and Beyond: An ROI Model. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Plant, R. A. and Ryan, R. J. (1994). Who is Evaluating Training? *Journal of European Industrial Training*, 18(5), 27-30, MCB University Press.

Popham, W. J. (1975). *Educational Evaluation*. Prentice Hall, Englewood Cliffs, NJ, USA.

Price, C., Hunt, J., Lee, M., and Ormsbey, A. (1992). A model-based approach to the automation of Failure Mode Effects Analysis for Design. *The Journal of Automobile Engineering*.

Provus, M. (1971). *Discrepancy Evaluation*. McCutchan, CA, USA.

Rae, L. (1991). *How to Measure Training Effectiveness*. Gower, Aldershot, Hants.

Raffe, D., Bundell, I., and Bibby, J. (1989). Ethics and Tactics: Issues Arising from an Educational Survey. In Burgess, R. G. (Ed.). *The Ethics of Educational Research*. Falmer Press, Lewes.

Ralphs, L. T. and Stephan, E. (1986). HRD in the Fortune 500. *Training and Development Journal*, 40, 69-74, CA, USA.

Reason, P. and Rowan, J. (Eds.) (1981). *Human Enquiry: A Sourcebook of New Paradigm Research*. John Wiley and Sons, Chichester.

Reid, M. A. and Barrington, H. (1994). *Training Interventions: Managing Employee Development*. Institute of Personnel and Development, London.

Reigeluth, C. M. (1983) *Instructional Design: What is It and Why is It?* In Reigeluth, C. M. (Ed.) (op.cit.)

Reigeluth, C. M. (Ed.) (1983). *Instructional-Design Theories and Models: An Overview of their Current Status*. Lawrence Erlbaum, NJ, USA.

Reigeluth, C. M. and Stein, F. S. (1983). The Elaboration Theory of Instruction. In Reigeluth, C.M. (Ed.) (op.cit.).

Rheingold, H. (1988). *They Have a Word for it*. Tarcher, Los Angeles, CA, USA.

Riechen, H. W. and Boruch, R. F. (Eds.) (1974). *Social Experimentation: A method for Planning and Evaluating Social Intervention*. Academic Press, New York, USA.

Rogers, E. M. (1983). *Diffusion of Innovations*. Free Press, New York, USA.

Rokeach, M. (1973). *The Nature of Human Values*. The Free Press, New York, USA.

Ross-Ashby, W. (1956). *An Introduction to Cybernetics*. Methuen

Rossi, P. H., and Freeman, H. E. (1993). *Evaluation: A Systematic Approach*. Sage Publications, London.

Rotter, J. B. (1972). An Introduction to Social Learning Theory. In Rotter, J. B., Chance, J. E., & Phares, E. J. (Eds.). *Applications of a Social Learning Theory of Personality*. Holt, Rinehart and Winston, New York, USA.

Ruddock, R. (1981). Evaluation: A Consideration of Principles and Methods. *Manchester Monographs*, 18, Manchester.

SAE (1994). *Potential Failure Mode and Effects Analysis in Design (Design FMEA) and Potential Failure Mode and Effects Analysis in Manufacturing and Assembly Processes (Process FMEA)*. Reference Manual SA J 1739. 1994-07. Society of Automotive Engineers, London.

Samaio, P. S. (1995). *Application of Structured Failure Mode and Effects Analysis in Organisational Change Implementation*. SAE Brazil. Technical Paper Series 952300. October 1995 (p.1)

Sathe, V. (1983). Implications of Corporate Culture: A Managers' Guide to Action, *Organizational Dynamics*, Autumn.

Scandura, J. M. (1983). Instructional Strategies based on Structural Learning Theory. In Reigeluth, C.M. (Ed.) (op.cit.).

Scandura, J. M. and Brainerd, C. J. (1978). *Structural/Process Models of Complex Human Behaviour. Sijthoff*, Leyden, Netherlands.

Scheffler, I. (1961). *Conditions of Knowledge - An Introduction to Espistomolgy and Education*. Scott Foresman.

Schein, E. H. (1985). *Organisational Culture and Leadership*. Jossey-Bass, New York, USA.

Scherkenbach, W. W. (1988). *The Deming Route to Quality and Productivity*. Mercury Press, Rockville, MD, USA.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82(4), 225 - 260.

Scholtes, P. (1988). *The Team Handbook*. Joiner Associates, Madison, WI, USA.

Scriven, M. (1991). *Evaluation Thesaurus*. Sage Publications, Newbury Park, CA, USA.

Scriven, M. S. (1996) Evaluation Ideologies. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (Eds.) (op.cit.).

Scriven, M. S. (1995). The Science of Valuing. In Shadish, W.R., Cook, T.D., and Leviton, L. C. (Eds.) (op.cit.).

Senge, P. M., Kleiner, A., Roberts, C., Ross, R. B., & Smith, B. J. (1994). *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization*. Doubleday, New York, USA.

Settle, T. J. (1987) in Hawthorne E. M. (1987). *Evaluating Employee Training Programs*. Quorum Books, New York, USA.

Shadish, W. R., Cook, T. D., and Leviton, L. C. (Eds.) (1995). *Foundations of Program Evaluation: Theories of Practice*. Sage Publications, London.

Shennen, F. and Lochhead, A. (1996). Forum for Beginners: Training Evaluation, *CA Magazine*, July.

Shewhart, W. A. (1931). *Economic Control of Manufactured Product*. D. Van Nostrand Company Inc., Princeton, NJ (USA).

Simpson, E. L. (1980). Adult Learning Theory: A State of the Art. In Hasker, H., Moore, J., and Simpson, E. L. (Eds.). *Adult Development and Approaches to Learning*. National Institute of Education, Washington, DC.

Skinner, B. F. (1938), *The Behaviour of Organisms: An experimental Analysis*. Prentice-Hall, Englewood Cliffs, NJ, USA.

Sloman, M. (1994). *A Handbook for Training Strategy*. Gower, Aldershot, Hampshire.

SMMT (1989). *Guidelines to Failure Mode and Effects Analysis*. The Society of Motor Manufacturers and Traders Limited, London.

Sneath, P. H. A., and Sokal, R. R. (1973). *Numerical Taxonomy*. Freeman, San Francisco, CA, USA.

Sockett, H. and Pring, R. (1971). Bloom's Taxonomy: A philosophical critique. *Cambridge Journal of Education* (1, 2), Cambridge.

Stake, R. E. (1967). The Countenance of Educational Evaluation. *Teachers College Record,* 68, 523-540. Englewood Cliffs, NJ, USA.

Stake, R. E. (1975). *Programme Evaluation, Particularly Responsive Evaluation*. Occasional paper No. 5, The Evaluation Center, Western Michigan University, Kalamazoo, MN, USA.

Stake, R. E. (1976). *Evaluating Educational Programmes*. Organisation for Economic Co-operational Development, USA.

Stake, R. E. (1981). Setting Standards for Educational Evaluators. *Evaluation News*, 2(7), 5-8.

Stake, R. E. (1996) Program Evaluation, Particularly Responsive Evaluation in Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (Eds.) (op.cit.).

Stammers, R. B. (1987). Training and the Acquisition of Knowledge and Skill. In Warr, P. *Psychology at Work*. Penguin, London.

Steinmetz, A. (1976a) The Discrepancy Evaluation Model. Measurement in Education, no. 1, 7 (pp. 1-7). In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (Eds.) (1996) (op.cit.).

Stratton, P., Munton, A. G., Hanks, H. G. I., Heard, D. H., and Davidson, C. (1988). Leeds Attributional Coding System Manual. LFRTC, Leeds.

Stufflebeam, D. L. (1966). A depth study of the evaluation requirement. *Theory into Practice*, 5, 121-34.

Stufflebeam, D. L. (1996) The CIPP Model for Program Evaluation in Madaus, G.F., Scriven, M. S., and Stufflebeam, D. L. (Eds.) (op.cit.).

Stufflebeam, D. L., & Madaus, G. F. (1996). The Standards for Evaluation of Educational Programs, Projects, and Materials: A Description and Summary. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (Eds.) (op.cit.).

Sudman, S. and Bradburn, N. M. (1982). *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass, Oxford.

Sugarman, B. (1998). The Learning Organisation: Implications for Training. In Brown, S. M., and Seidner, C. J. (Eds.) (op.cit.).

Syer, J. & Connolly, C. (1987). *Sporting Body, Sporting Mind.* Simon and Schuster, London.

Syer, J. (1986). *Team Spirit.* Kingswood Press, London.

Taguchi, G. (1986). *Introduction to Quality Engineering*. Quality Resources, White Plains, New York, USA

Tannenbaum, S.I., & Yukl, G. (1992). Training and Development in Work Organisations. *Annual Review of Psychology,* 43, 399-441.

Taylor, F. W. (1919). *Shop Management*. Harper and Brothers, New York, NY, USA.

Thackwray, R. (1997). *Effective Evaluation of Training and Development in Higher Education*. Kogan Page, London.

The Joint Committee on Standards for Educational Education (1981). *Standards for the Evaluation of Educational Programs, Projects and Materials*. Sage Publications, Newbury Park, CA, USA.

The Training Agency (1989). *Training in Britain, A Set of Research Reports.* HMSO, London.

Thomas J. (1985). Force Field Analysis: A New Way to Evaluate Your Strategy, *Long Range Planning*, 18(6), 54-59.

Thorndike, E. L. (1913), *Educational Psychology: The Psychology of Learning. Vol. 2.* Teachers College, New York, USA.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review,* 8, 247-261.

Tichy, N. M., and Sherman, S. (1995). *Control Your Destiny or Someone Else Will.* Harper Collins, London.

Tolman, E. C. (1932). *Purposive Behavior in Animals and Men*. Appleton-Century-Crofts, New York. Reprinted by University of California Press, 1949.

Tyler, R.W. (1950) *Basic Principles of Curriculum and Instruction*, Chicago, Ill., University of Chicago Press.

USA Department of Defence (1980). *MIL-STD-1629A: Procedures for performing a Failure Mode, Effects and Criticality Analysis (Nov)*.

Wagner, R. and Weigand, R. (1994). Evaluating an Outdoor-Based Training Program. In Kirkpatrick, D. L. (Ed.) (op.cit.).

Walker, J. W. (1992). *Human Resource Strategy*. McGraw-Hill, New York, NY, USA.

Weiner, B. (Ed.) (1974). *Achievement Motivation and Attribution Theory*. General Learning Press, Morristown, NJ, USA.

Weiss, C. H. (1986). The Stakeholder Approach to Evaluation: Origins and Promise. In House E. R. (Ed.) (op.cit.).

Wexley, K. N. (1984). Personnel Training. *Annual Review of Psychology*, 35, 519-551.

Wexley, K. N., & Latham, G. P. (1981). *Developing and Training Human Resources in Organizations*. Scott Foresman, Glenview, Ill., USA.

Worthen, B. R. and Sanders, J. R. (1973). *Educational Evaluation: Theory and Practice*. Charles A. Jones, Worthington, Ohio, USA.

Woodcock, M. and Francis, D. (1990). *Unblocking Your Organisation*. Gower, Aldershot, Hants.

Zairi, M. (1991). *Total Quality Management for Engineers*. Woodhead Publishing Ltd.

Appendices

Appendix A Theoretical Bases for Training Design

The following theoretical perspectives are identified by Reigeluth (1983) as being significant to the design of training and therefore have particular relevance to this study.

(I) Prescriptive Model of Instruction

Gagne and Briggs (1979) distinguish five different types of learning (see chapter 1: section 1.2.4) and focuses on nine training design events of; (i) gaining attention; (ii) informing learners of the objective; (iii) stimulating recall of prior learning; (iv) presenting the stimuli; (v) providing learning guidance; (vi) eliciting performance; (vii) providing feedback; (viii) assessing performance; and (ix) enhancing retention and transfer.

(II) Behavioural Approach to Instructional Prescription

Gropper (1974, 1975, 1983) emphasises the role of practice. The theory is grounded in behaviouralist Stimulus-Response concepts of learning and central to the theory is the learner's practice of designated responses in the presence of criterion stimulus. As learners must be able to distinguish between stimulus and related responses, the theory requires practice in associating the criterion stimulus and criterion response. Furthermore, to learn a total skill, students must be given the opportunity to practice chaining total series of S-R associations that make up the skill.

The theory employs four disciplines of (i) discrimination; (ii) generalisation; (iii) association; and (iv) chaining, as building blocks of all types of objectives.

(III) Algo-heuristic Theory of Instruction

Landa (1983) seeks to simplify complex skills into elementary cognitive (or motor) operations that can be executed by learners in the course of learning and performance. These elementary cognitive operations are combined into either algorithmic or heuristic processes which underpin complex intellectual tasks. An algorithmic process is a set of parallel and serial elementary operations assembled to solve all problems of a certain type. Heuristic processes consist of a series of non-elementary operations, or elementary operations which are not performed in a uniform way. Landa (1983) acknowledges that not all intellectual activity can be described using algorithms. It can be neither possible or practical, particularly in an instructional setting to attempt to prescribe algorithms for complex operations.

The algo-heuristic theory, therefore, deals primarily breaking down cognitive (or motor) skills into their elementary components (elementary from the viewpoint of the learner) and formulating them into step by step processes to facilitate understanding and learning by providing simple explicit models of skill performance.

Within the engineering quality improvement programme, algorithms are used extensively, in the form of flowcharts, to conceptualise many of the quality methodologies (i.e. Team Oriented problem solving; Failure mode and Effects analysis) to facilitate trainee learning.

(IV) Structural Learning Theory

Scandura (1983); and Scandura and Brainerd (1978) emphasise the role of rule learning. Scandura describes the theory as a series of steps; (i) identifying educational goals in terms of what the learner will be able to do after training; (ii) identifying the associated prototypic processes - how the learner is to perform tasks associated with educational goals; (iii) characterisation of the individual learners in terms of what they know prior to training; and (iv) the ongoing and goal-directed interactional process between trainer and trainee.

Scandura (1983) provides a general method of analysis for establishing prototypic processes, or sets of rules, based on the educational goals. He calls this structural analysis which is essentially a process of deriving simple rules which determine the execution of a task. The lower the level of prior knowledge and understanding of trainees, the simpler the rules should be. The rules are then sequenced into training. Scandura advocates using expository (telling) or discovery methods of instruction to teach the rules.

Scandura's Structured Learning Theory provides for defining logical sequences or rules to facilitate learning. The theory is problem oriented and its emphasis on logic is made even more apparent given Scandura's mathematical background.

(V) Cognitive Theory of Inquiry Teaching

Collins and Stevens (1983) method is concerned with how to train cognitive strategies by enabling the trainee to discover all of the required factors in a theory and make predictions from it. Collins and Stevens identify strategies which good trainers employ in tutorial dialogues which aim to; teach causal relationships (i.e. theories and models) in a topic area which require trainers to diagnose any trainee misconceptions and overcome them; and enable trainees to derive new theories in a domain from a series of examples or cases.

The theory is made up of three components; (i) the goals of the training; (ii) the strategies used to achieve the goals; and (iii) the control structure which governs the trainer/trainee dialogue. Collins and Stevens identify two top level goals of; teaching trainees particular rules or theories; and teaching trainees how to derive rules or theories.

From analysing lessons of a variety of trainers, Collins and Stevens identify ten strategies for inquiry teaching; (i) selecting positive and negative exemplars; (ii) varying cases systematically; (iii) selecting counter-examples; (iv) generating hypothetical cases; (v) forming hypotheses; (vi) testing hypotheses; (vii) considering alternative predictions; (viii) entrapping students; (ix) tracing consequences to a contradiction; and (x) questioning authority. Guidance on the selection and use of the ten strategies is given to facilitate enquiry training.

Thirdly, the dialogue control structure enables the trainer to allocate time between different goals to optimum effect. Collins and Stevens describe the control structure as consisting of; a set of strategies for selecting cases with respect to the top level goals; a student model; an agenda; and a set of priority rules for adding goals.

Whilst Reigeluth and Patrick classify *Cognitive Theory of Inquiry Teaching* as a training (instructional) design theory, its narrow range of application, with regard to different types of learners (see Honey and Mumford, 1992), makes it more specific and therefore more accurately (and helpfully) classified as a training delivery method.

(VI) Component Display Theory

Merrill (1983) provides a framework for integrating instructional principles of Gagné and others. Types of learning, training objectives, test items, and instructional presentations are mapped to each other in an attempt to ensure that presentations and tests are adequate and consistent with each other. Merrill defines ten types of learning using a Performance - Content matrix from which generic training objectives are derived. In addition, Merrill offers several types of instructional presentation which training designers can sequence into a training programme.

(VII) Elaboration Theory of Instruction

Reigeluth and Stein (1983) prescribes training methods which consider many related topics in an overall training programme. Reigeluth and Stein describe the model in terms of four problem areas; (i) selection; (ii) sequencing; (iii) synthesising; and (iv) summarising. The theory prescribes commencing training with an overview that teaches a few general, simple, and fundamental ideas, with the remainder of the training considering progressively more detailed ideas, which elaborate on earlier ones. The theory also prescribes the systematic use of review and synthesis as part of the simple to complex instructional design.

The theory is derived largely from the work of; Gagne (1977), Bergan (1980) with respect to learning prerequisites and hierarchies respectively; Gropper (1974). Landa (1974) and others with respect to procedural relationships of instruction; Ausubel (1968), Bruner (1960) and others with respect to sequencing. According to the theory's authors, Elaboration theory integrate all of this previous research.

The general principles of elaboration theory are employed in the design of the engineers' training programme; the modular structure of the programme commences with the Programme Foundation model; which provides a conceptual overview of the programmes quality methodologies and their general relationships to each other. Within each module, the design of training specifies an introduction and overview of the content of the module, progresses with more detailed attention of each aspect of the content and concludes the module with a summary of the overall methodology, usually in the form of an application to a case study

(VIII) Motivational Design of Instruction

Keller (1983) synthesises theoretical notions of motivation into principles relevant to training design. Keller identifies four basic categories of motivational conditions; interest; relevance; expectancy; and satisfaction. *Interest* refers to whether the trainees curiosity is raised during training and whether this arousal is maintained over time. *Relevance* refers to the trainee's perception of personal need satisfaction in relation to the training. *Expectancy* refers to the trainee's perceived likelihood of success and the extent to which success is under the trainee's control. *Satisfaction* is concerned with extrinsic rewards and intrinsic motivation and whether these are compatible with the trainee's anticipation.

Keller offers a model for designing motivating instruction which provides a basis for integrating several strategies for increasing motivation in terms of conditions, methods and outcomes of the training. For example, with regard to *interest*, strategies suggested are; (i) using novel conflictual, and procedural events to engage trainees' interest and increase curiosity; (ii) using anecdotes and other devices to inject a personal emotional element into what would otherwise be purely intellectual material; and (iii) giving trainees opportunity to learn more about aspects they already know, but include perspectives which are unexpected and unfamiliar.

Keller has drawn on much of the motivational research in formulating his strategies (e.g. Berlyne, 1965 with respect to curiosity and arousal; Maslow, 1954 and McClelland, 1976 with respect to personal needs; Feather, 1975 and Rokeach, 1973 with respect to beliefs and attitudes; Rotter 1972 with respect to locus of control; Weiner, 1974 with respect to attribution theory; and Bandura, 1977 with respect to self efficacy).

Motivational design of instruction, more so than any of the other theories of training design, pulls together much of the research on motivation. Motivation of the learner is widely acknowledged in the literature as being significant to learning and the transfer of training into workplace behaviours. As will be discussed in the next chapter, as part of the review of the CIPP model of evaluation, the significance of the motivation of trainees to the effectiveness of training makes it an important aspect to consider in the evaluation of training.

Appendix B Conceptual Models of Evaluation

The following conceptualisations of evaluation have been drawn from the literature and are included here to provide a descriptive overview of their purpose, framework and methodology.

(I) Discrepancy Evaluation Model

Originally conceived by Provus in 1969 and developed by Provus (1971) and Steinmetz (1976), discrepancy evaluation is a comparison of performance (P) against standards (S) where standards are descriptions of the qualities or characteristics of the evaluand. The difference between the two is the discrepancy (D) information. Evaluation, therefore, is making judgements about the worth or adequacy of an evaluand based upon discrepancy information.

With regard to the *purpose* of evaluation, discrepancy evaluation is concerned with measuring performance against performance criteria expressed in terms of a specification. Steinmetz (1976) proposes the purpose as programme improvement by making the SPD cycle explicit and public.

With regard to evaluation *methodology*, discrepancy evaluation has a three stage methodology, although these are not necessarily discretely sequential and can be conceived as a 3 stage continuous cycle. The first stage is the establishment of the standard, which is largely undertaken as a consultative process between the evaluator and the client(s). During this stage qualities and characteristics of the evaluand are identified and ideal performance levels, or performance criteria, are agreed. The role, with respect to data collection and judgement are also agreed between the evaluator and the client.

The second stage of the methodology is concerned with the establishment and application of measures of performance with respect to the specifications. This may take the form of attribute (OK / not OK) measures or variable measures. The third stage, is the comparison of performance to specifications and the description of nature and magnitude of discrepancies.

With regard to the evaluation *roles*, Steinmetz positions the evaluator as a facilitator to the evaluation, soliciting information from the client to help the client formulate the specification and identify what kinds of information would constitute performance in terms of the specification. With regard to collecting performance data and the assessment of discrepancies, Steinmetz proposes that this would normally be the role of the evaluator, however the nature of the performance information and the subsequent discrepancy decision may determine the need for the client to collect the data. Steinmetz gives an example of ride stability of a motorcycle, where stability performance is a judgement value which is subject to the likes and dislikes of the would be rider, in which case the evaluator would not be in a position to decide what the client prefers.

(II) Context, Input, Process, Product (CIPP) Model

Conceived by Stufflebeam (1966) as a decision oriented model of evaluation, CIPP incorporates evaluation of a programme's context, inputs, processes and products.

Context evaluation is concerned with defining the target audience and their needs, opportunities for addressing the needs, and to judge whether proposed programme objectives are sufficiently responsive to their needs. Context evaluation is concerned with the programme before it is launched. The information generated is used for planning needed changes and for providing a basis for judging outcomes.

In training terms, context evaluation shares similar aims to training needs analysis. Within the training literature (i.e. Patrick, 1992; Goldstein et al, 1986; Reid & Barrington, 1994) however, training needs analysis is distinguished from, and beyond the scope of, training evaluation.

Input evaluation is concerned with the assessment of programme implementation strategy; scheduling, procedures and administration. It assesses the system's capability to meet the defined needs. Input evaluation is used for selecting sources of support and procedural design.

Within a training context, participant's receiving training at the right time; a critical factor to the transfer of training (Broad & Newstrum, 1992) would be the subject of Input evaluation.

Process evaluation has the objective of identifying, or predicting, defects in the procedural design or it's implementation. It provides feedback on the execution of the evaluand for improvement and development of process controls. It also provides source information for the interpretation of product (outcome) evaluation.

Product evaluation collects descriptions and judgements of outcomes (intended and unintended, positive and negative) and relates them to programme objectives and to context, input, and process information. It interprets the programme's worth and merit and is used for deciding whether to continue, terminate, modify, or refocus the programme.

Stufflebeam advocates using any combination of the CIPP evaluations to evaluate a programme. To guide the implementation of the evaluation, he calls for the preparation of preliminary plans and warns that the dynamic and interactive qualities of many evaluations may render the technical plans for data collection and analysis inappropriate. The evaluation design must therefore be viewed as a process, with evaluation goals and procedures sketched in advance and subsequently reviewed, revised, expanded and operationalised. So as not to undermine the confidence of the client of the evaluation, the evaluator should involve the client in the ongoing design / redesign process.

More recently and within a training context, Easterby-Smith (1994) has elaborated Stufflebeam's model into a five element framework for evaluation; context, administration; inputs; process; and outcomes. Little empirical support is offered in the literature for this conceptualisation in a training context.

(III) Illuminative Evaluation

Illuminative evaluation (Parlett and Hamilton, 1972) was developed in the course of studying small-scale educational programs. Parlett and Dearden (1977) describe Illuminative evaluation as:

> *"The basic emphasis of this approach is on interpreting, in each study, a variety of educational practices, participants' experiences, institutional procedures, and management problems in ways that are recognisable and useful to those for whom the study is made. The illuminative evaluator contributes to decision making by providing information, comment, and analysis designed to increase knowledge and understanding of the programme under review. Illuminative evaluation is characterised by a flexible methodology that capitalises on available resources and opportunities, and draws upon different techniques to fit the total circumstances of each study."*

Illuminative evaluation overcomes evaluand participant and stakeholder resistance, which stems from suspicion of the evaluator, or the political sensitivity and implications of the evaluation findings (Parlett, 1981). Illuminative evaluations, like 4th Generation (page 300), are not designed in advance. They are exploratory, with the evaluation questions being identified as the study progresses and as critical issues of the programmes emerge. The evaluation methodology commences with the identification of relevant participants and stakeholders and the collecting of data through non-directive methods, such as unstructured interviewing. Following analysis of data to identify issues of interest, a cycle of data collection / data analysis is pursued to investigate identified issues and to identify further issues. Interpretations of value and worth are made by participants and stakeholders.

Brandes (1985) employed illuminative evaluation to investigate and evaluate an alcohol education programme and concluded that the methodology provided a flexible evaluation framework which incorporated the needs of the programme's stakeholder whilst maintaining the credibility of the evaluation.

(IV) Responsive Evaluation

Developed by Stake (1975) from his early writings (i.e. Stake, 1967), where he emphasised the use of multiple sources of information for the evaluation of education - an approach he termed 'Countenance Model of Evaluation', responsive evaluation is, as the name suggests, responding to events that are noticed about the evaluand and are of interest to evaluand stakeholders. There is no predetermined evaluation design and Stake offers this as a contrast to what he calls 'preordinate evaluation'. The emphasis is on rich qualitative data and personal accounts of events.

With regard to the *purpose* of evaluation, Stake (1996) argues that there are many purposes to evaluation including, amongst others; documenting events, recording student changes; detecting institutional vitality; placing blame for trouble; aiding administrative decision-making; facilitating corrective action; and increasing understanding of learning and teaching. As each of these purposes relate directly or indirectly to the values of a programme, Stake argues they can be considered legitimate purposes for evaluation.

The purpose of responsive evaluation, therefore, has to be decided by the evaluator and based on the particular situation and giving careful attention to the reasons the evaluation was commissioned. The rationale for Stake's position is his argument that the 'pay-off' in terms of outcomes may be "diffuse, long-delayed or beyond the scrutiny of the evaluator .... and therefore the evaluator should not presume that only measurable outcomes testify to the worth of programme" (Stake, 1996: pp 294).

With regard to evaluation *methodology*, Stake does not prescribe a methodology for responsive evaluation, preferring to advocate the formulation of a plan of observation and feedback. The plan is structured around the programme issues, identified as part of the purpose definition process, and the issues may change as the programme and evaluation progress. The emphasis is on providing portrayals of what is happening with a programme using scripts, log-books, exhibits, and tape recordings. Through the feedback communication process the evaluation is re-focused to follow emerging issues.

Stake acknowledges the highly subjective nature of this approach, however argues that the validity of the evaluation is measured by the degree of endorsement of large numbers of audience-significant people. Stake positions responsive evaluation as an and / or alternative to what he describes as preordinate evaluation where goals and objectives set the criteria for evaluation.

With regard to the evaluation *roles*, Stake positions the evaluator as the planner/investigator/ reporter. The responsibility for undertaking the responsive evaluator rests with the evaluator.

(V) Connoisseurship Model of Evaluation

Eisner (1976) proposed a model which is directly judgmental and influenced greatly by his persuasion to teaching in the arts. The underlying principles of the model are that of connoisseurship and criticism, where connoisseurship is an awareness and understanding of educational experiences which comes from a high level of expertise, and criticism is the description, interpretation and appraisal of the experiences.

With regard to the *purpose* of evaluation, Eisner argues that the ultimate objective of educational evaluation is the improvement of the quality of educational life students lead. Eisner regards the connoisseurship model as a way of broadening the base of educational evaluation along side scientifically oriented approaches to evaluation.

With regard to evaluation *methodology*, the evaluator must become emerged in the evaluand to experience and observe its characteristics. Information is collected to describe phenomena which is interpreted and used to appraise the phenomena. The evaluator (or critic) uses what he or she sees and interprets in order to arrive at some conclusions about the character of educational improvement or practice. Criteria for appraisal is not explicit, it relies of the judgement of the evaluator.

Eisner proposes that the validity of the evaluation comes from structural collaboration; a process that seeks to validate or support conclusions by demonstrating how a variety of facts or conditions within a phenomena support the conclusions drawn. Eisner cites jurisprudence as an activity which bases validity on structural collaboration where two barristers present the facts in a case to prove / disprove guilt and a jury concur or fail to concur that the evidence is sufficiently coherent and cohesive.

With regard to the evaluation *roles*, by the nature of model, the evaluation can only be undertaken by persons who have a high level of subject and process expertise. As the evaluator is required to formulate criticisms of the evaluand, independence from the evaluand is necessary.

(VI) 4th Generation Evaluation

Fourth Generation Evaluation was conceived by Guba & Lincoln (1989). Their rational was that evaluation research had evolved through 3 generations from the conventional (Inquiry) paradigm to constructivist Inquiry paradigm. They advocate two fundamental tenants of evaluation; *responsive focusing* - the evaluation questions are determined from information collected from stakeholders (those with a vested interest in the evaluand); and *constructivist methodology* - making inquiry using 'hermeneutic dialectic' processes, that is to say, cyclical processes, which with each cycle generate more information and more questions.

Guba & Lincoln argue that research inquiry methods cannot be established beforehand as evaluation is an emergent process. Its focus (or foci) are dependent on inputs from stakeholders and its activities are 'serially contingent' (p.263).

Responsive focusing requires that research agendas are developed from interviews with selected stakeholders, and constructivist methodology requires that stakeholders provide data to give themes and issues to be pursued in the evaluation. Guba and Lincoln offer operational guidelines for undertaking 4th Generation Evaluation. These are given in Table B-I.

Step	Description
1. Contracting	Initiate the contract with the client or sponsor of the evaluation - those who are legally / fiscally in a position to contract the evaluation.
2. Organising	Select and train the team of evaluators. Make entree arrangements, logistical arrangements and assess the local political factors.
3. Identifying Stakeholders	Identify the agents, beneficiaries, victims of the evaluation - the persons who are put at risk by the evaluation (Lincoln & Guba,1989). As the evaluation progresses, new stakeholders may be identified, requiring continuing stakeholder search strategies. The stakeholder membership and participation agreements are formalised by being included in the conditions of the contract with the sponsor or client.
4. Developing within-group joint constructions	The joint evaluation construction of the group is established using the hermeneutic circle approach. Stakeholders, in turn, respond to the established constructions, and in the process reconstruct. This cycle is repeated until agreed upon constructions are established within the stakeholder group.
5. Enlarging joint stakeholder constructions through new information / increased sophistication	Other information, from outside the stakeholder group, is systematically introduced to "inform the constructions further and raise them to a higher level of sophistication". This 'new' information comes from documents and records relevant to the evaluand, observation of the evaluand, professional literature (the body of knowledge that exists in journals), other stakeholder circles (stakeholders excluded from the stakeholder group), and the evaluator's etic (outsider) construction - the evaluator's priori knowledge and opinion.
6. Sorting out unresolved claims, concerns, and issues	Identify claims, concerns, and issues surfaced during steps 5 and 6 and were resolved by consensus in the group for future settings. These are set aside as case report components for this purpose.
7. Prioritising unresolved items	Not all claims, concerns, and issues which arose in stages 4 & 5 would have been resolved and therefore, a participatory prioritising process is determined. The items are submit to prioritisation, with low priority items being left unresolved as may be the case.
8. Collecting information / adding sophistication	To resolve all / high priority items, further information is collected and added to the constructions, thereby collaboratively reconstructing them.
9. Preparing agenda for negotiation	The unresolved claims, concerns and issues, and the collection of information relating to them, requires the evaluator to prepare an agenda for stakeholder negotiation to clarify and agree what the data collected will mean (in a positivist paradigm this would not be an issue - data are facts, and facts are reality). This step is analogous to drawing up a set of conclusions and recommendations
10. Carrying out the negotiation	The negotiation is carried out via the hermeneutic circle (see step 4).
11. Reporting	The 4th generation evaluation report is necessarily longer than that of conventional evaluation; the 'facts' are described in stakeholders terms giving the reader an understanding of how the stakeholders make sense of the facts.
12. Recycling	4th Generation Evaluations tend to raise more questions than they answer - they are divergent. The unresolved claims, concerns, and issues provide further evaluation questions. Also, the constructions (results of the evaluation) may hold for a finite period of time. New information may be made available thereby reopening the evaluation. According to Guba and Lincoln, "Fourth generation evaluations never stop, they merely pause" (p. 226).

Table B-I: The Flow of 4th Generation Evaluation - Guba & Lincoln (1989)

(VII) Utility Analysis

Utility Analysis is a powerful tool for expressing the outcomes of personnel programmes in terms of dollars (Boudreau, 1983; Cascio, 1980, 1987; Cascio and Ramos, 1986; Schmidt, Hunter, McKenzie, and Muldrow, 1979; Schmidt and Hunter, 1981; 1983).

Phillips (1994) advocates the use of 'Return on Investment' for measuring the effectiveness of "Human Resource development" programs, explaining that the HRD function (Education and Training) cannot continue to operate "in a world without accountability".

Within the context of the evaluation of training, little attention has been given to the wide range of conceptualisations of evaluation which feature in the arena of the evaluation of education. This may be due, to some extent, to the distinction which has been drawn between education and training. Popular writers on the evaluation of training, particularly in a commercial context (i.e. Phillips, 1991; Patrick; 1992; and Bramley, 1997) have concentrated on evaluation conceptualisations based largely on Kirkpatrick.

One aspect of evaluation of education which has been adopted in the context of the evaluation of training however is the notion of evaluand stakeholders.

(VIII) Stakeholder Analysis

Stakeholder Analysis is a research tool that recognises that for any organisational change programme, there are interested parties who affect, or are affected, by the programme (Burgoyne, 1992)

The Stakeholder approach was developed by the American National Institute of Education (Weiss, 1986) to address concerns about the perceived lack of fit between evaluation and the socio-political context of an evaluand. Weiss cited five areas of concern with existing approaches to evaluation within the field of education. These were:

1) That the studies were <u>narrow</u> as they considered a limited number of issues which were selected for examination at the outset. The issues selected did not always turn out to be those which were crucial to the people most concerned with the program.

2) That evaluators established <u>unrealistic</u> expectations by selecting measurement systems (scales and instruments) which were not sufficiently sensitive to detect likely realistic changes and subsequently doom the program to failure. An analogy of this would be using a road-side weigh-bridge to measure changes in a person's weight on a diet programme. The measurement scale on the weigh-bridge is in tonnes and yet changes to a person's weight would be in Kilograms, or even Grams.

3) The evaluation by outcome data collected and analysed are <u>irrelevant</u> and unresponsive to the real needs of the people involved in the program, who may, for example, be interested in why people drop out of the program, why the program fails to attract participants and other operational concerns.

4) The evaluations are <u>unfair</u> in that they address questions which are important to the sponsoring client and not questions which are important to others who are involved and often affected by the program.

5) Probably the most significant, in my view, that evaluation findings are <u>unused</u>. Weiss observed that evaluation results were seldom used to affect decisions about a programme.

The stakeholder approach was designed by the American National Institute of Education to increase the use of evaluation results for decision making and to bring a wider variety of people into active participation in the evaluation process, thereby addressing the issues outlined above.

The purpose of stakeholder analysis is to elicit the views of stakeholders as the basis for the evaluation. In the sense of Weiss (1986), stakeholder analysis is used to identify the values, as defined by those people who affect and are affected by the training, against which training is evaluated and is therefore an aspect of the planning stages of an evaluation. The value criteria for the evaluation is drawn from the stakeholder analysis. Burgoyne (1992; 1994) emphasises that stakeholders' may have different perceptions and constructs from their experience of a programme and these provide the data for an evaluation. Stakeholder analysis provides a method where these different constructions and causal beliefs can be elicited and understood for the purposes of the evaluation.

Appendix C Programme Content Overview

The programme content comprises a range of technical (quality) methods and tools' and behavioural (people) skills' intended to improve the engineering process. The primary technical methods are: disciplined problem solving, process management, problem prevention, experimentation, quality engineering, and customer focussed[1] engineering. The primary technical methods are supplemented by basic quality tools which also provide a foundation for many of the technical methods. The behavioural skills are; team building; communication; implementation; and innovation.

The following provides a brief introduction to the technical and people skills content of the programme, which summarises the training materials consisting of over 500,000 words, plus numerous graphics.

(I) Basic Quality Tools

The basic quality tools are statistical and graphical tools for data collection and analysis. The set of eight tools comprise of what are commonly known as Ishikawa's (1982) Seven Tools; graphs, histograms (Guerry, 1833), cause and effect diagrams (Ishikawa, 1943), check sheets, Pareto Diagrams (Pareto, 1896), Control Charts (Shewhart, 1931) and scatter diagrams. In addition to these seven, flow diagrams were added by the programme development team. A brief description of each tool, together with its purpose and application are given in Table C-II and Table C-III.

The basic quality tools are data collection, analysis and communication tools which are applied within the frameworks of all quality methods in the programme. In this sense they are fundamental to the understanding and application of the quality methods.

[1] deliberately mis-spelt

Description	Purpose and Application
Graphs provide a way to display data in a pictorial format. They can take many forms including line graphs, bar graphs, box and whisker plots and pie charts. The reader is probably familiar with all of these with, perhaps, the exception of box and whisker plots. These are used to show the mean (box) and range (whisker) of the data.	Graphs, if applied properly, can be an effective tool for analysing data and for communication.
Check sheets provide a means to record data in pre-identified categories or as a planning tool for an event. Check sheets are normally designed for specific types of data collection and the format of the check sheet provides the recorder with a pictorial representation of the data showing frequency and/or location.	Three examples of check sheets given are: i) Recording check sheet, or tally chart, to collect data in pre-defined categories, ii) Check-list check sheet, to ensure all tasks of a project are performed, and iii) Location sheet, to identify where a particular activity is occurring, or the incidence of a particular activity at a location.
Pareto diagrams provide a means to analyse data. Collected data categories are represented as bars, with the height of each bar determined by the relative magnitude of the category to which it relates. The bars are drawn onto the diagram in descending order of height (magnitude) from left to right.	They are used to show the relative magnitudes of categories of data to determine priorities (e.g. the defect data collected on rejected spark-plugs at the end of a process may be categorised into cracked insulator, printing defect, missing cap, missing core electrode, failed functional test and damaged cap electrode). The occurrence of each category is scaled and plotted onto the diagram, with the most frequently occurring defect on the left and the least occurring on the right. This display of data enables a problem solving team to quickly identify the highest, or biggest cause of rejected plugs. The tool may be used further by the team to analyse a particular category (i.e. for printing defects, the collection and analysis of more data on the types of printing defect). In addition to this analytical aspect, Pareto diagrams provide a means of communication. A Lorenz line can also be plotted on the diagram, which displays the cumulative percentage of the categories.
Histograms are also a form of bar chart used to collect and analyse data. However, unlike Pareto diagrams, the bars of a histogram represent a value or range of values (classes) and the height of the bar represents its frequency of occurrence.	Histograms are used to display the range and pattern of variability in a process during the time frame the data is collected. E.g. a production process producing differential cases will have a lathe, or turning operation to machine the two bearing journals of the case. If the machined diameters of one of the bearing journals were plotted onto a histogram, the histogram would show a pattern of the variability of the process. From this picture of the process an operator or engineers could gain knowledge for process improvement. Two peaks, for example, would indicate 2 sets of distributions. Further investigation would reveal the causes leading in elimination of the cause and hence process improvement. Histograms can be used alone, or often, as in the case of my industry, in conjunction with control charts which are described later.

Table C-II: Basic Quality Tools

Description	Purpose and Application
Flowcharts display pictorially the sequence of events or activities in a process. They represent a detailed model of the process. In the case of a manufacturing assembly operation, the flowcharts shows the flow of the part through the operations of the process. Symbols are used to represent different type of activities (machining, storage, movement).	Flowcharts are primarily used for communication, although they can be used for analysis of a process - e.g. comparing the actual process with the one that has been specified. Three types of flowchart are explored; Process Flowchart to represent process operations and events of a process linked together using arrows depicting the flow of activities, Organisational flowchart to represent the flow of tasks and responsibilities through the organisation, time-based flowchart which not only represent the sequence, but the time taken by and time lapsed between tasks or operations.
Cause and effect diagrams display an effect linked by lines to its possible causes. They are also referred to as Ishikawa or fishbone diagrams. They can be used by an individual or a group to brainstorm possible causes to a problem, where the effect is written in a box with a horizontal line extending across the paper. Possible causes are drawn as branches from this line. Often causes will themselves have causes, which are written against twigs from the branches.	Three types of cause and effect diagrams are considered; cause enumeration where all possible causes are thought of, grouped by affinity and constructed in their affinity groups onto the cause and effect diagram, cause dispersion - where possible causes are thought of against cause categories (i.e. production process causes can fall into the 5 categories of Material, Equipment, People, Environment, Method), and production process type flowchart where the possible causes of variation (effect) are brainstormed against each stage of the process.
Scatter diagrams are graphical representations of the relationship between two variables. Points are plotted on the 2 axis of the graph (i.e. voltage against age of battery) and the relationship between one variable (voltage) and the other variable (age) can be analysed.	The type and strength of the relationships between the two variables can be analysed in terms of positive and negative type (angle of slope of best fit line) and how closely together the points fall on the graph. Where points are close together they are said to have a strong relationship and where they are spread all over the graph they are said to have no relationship.
Control charts are a type of line graph with a specific application to processes. The x axis is used to represent time and the y axis to represent measurements. Conceived by Shewhart (1931), the charts are used real-time to identify when a process has gone out of statistical control. The control chart has a centre line and control limit lines, all of which are calculated from previous y axis data.	The charts are used to record and analyse the performance of a process. Using a set of rules which identify non-random patterns of variation, an operator can quickly identify when a process has gone out of statistical control and take actions to investigate the cause. The main types of control charts considered are for variable data (e.g. X-bar, R) and for attribute data (e.g. p, np).

Table C-III: Basic Quality Tools (continued)

(II) Team Oriented Problem Solving

Team Oriented Problem Solving is a structured methodology for solving problems. Kepner and Tregoe (1981) define a problem as a deviation of a system's performance from its expected level. The Team Oriented Problem Solving Process was developed by the company as a standard corporate approach to problem solving, providing both a methodology for the resolution of problems and a common reporting format. The problem solving strategy comprises two elements; a) a rational process; and b) process facilitation techniques.

a) Rational Process: The problem solving approach is organised into an 8 discipline approach (Figure C-I). It is designed for application by problem solving teams, with two of the eight steps (1 & 8) placing emphasis on the team approach. Although the process advocates a team approach, it can be applied by an individual.

Eight Disciplines

1) Establish the Team:
Identify a small number of knowledgeable people with the skills and expertise to solve the problem and prevent its recurrence. A senior company champion should also be identified to provide management support for the team's actions.
2) Describe the Problem:
Specify the customer problem in quantified terms.
3) Implement Interim Containment Action:
Define and implement actions to keep the effects of the problem from the customer. Verify the effectiveness of the action with the customer.
4) Define and Verify Root Cause:
Identify all potential causes of the problem and by testing against the problem description and physical testing, isolate the root cause of the problem.
5) Choose and Verify Permanent Corrective Action:
Develop corrective actions that will provide robust solutions to the problem, and through test programmes using designed experimentation (where appropriate), verify that they will resolve the customer problem and not cause undesirable side effects. Anticipate potential failures associated with the chosen corrective, assess risks and plan contingency actions as necessary.
6) Implement Permanent Corrective Actions:
Establish plans and implement the permanent corrective action(s). Establish process controls and monitor effectiveness.
7) Prevent Recurrence:
Modify management systems, operating systems, practices and procedures to eliminate the possibility of this or a similar type of problem recurring.
8) Congratulate the Team:
Complete all aspects of the task and recognise team members' contributions to the problem solving efforts.

Figure C-I: Disciplines of Team Oriented Problem Solving

Its application is not restricted to problems of a technical or engineering, nature, although this is its primary purpose within the context of the programme, and can be applied to any situation where a problem is identified and the cause is not known and needs to be known. The eight disciplines provide a complete process for defining and solving a problem. The process offers the added benefit of extending beyond the resolution of a specific problem to preventing further problems of a similar nature occurring (step 7).

b) Process facilitation techniques: The second element recognises the need to provide teams with techniques to facilitate the problem solving approach. These are referred to as 'process helps' (Figure C-II), and intended for application in a general as well as a problem solving context.

Concerns Analysis:
A Methodology to prioritise concerns and to identify whether a past, present or future-oriented approach is required to address them, necessitating a problem solving, decision making, or problem prevention methodology respectively. The key elements of concerns analysis are to analyse the situation, subdivide complex issues into manageable elements, allocate priority on the basis of the seriousness, urgency or potential growth, and to identify which type of approach is appropriate.

Problem Solving:
A structured, data-driven methodology to identify the cause of a problem. It comprises four sub-processes; to describe the problem and develop its profile in quantified terms of identity, location, timing and magnitude; to postulate a range of possible causes; to select likely cause by analysis; and to verify by physical test that the true (root) cause has been identified.

Decision Making:
A methodology comprising seven steps; identifying the end result to be achieved by the decision; listing the decision criteria against which the decision will be judged in categories of "givens" (mandatory) and "wants" (desirable); deciding on the relative importance of the "wants"; identifying options to achieve the end result; evaluating the options against the criteria and prioritising the choices by calculating a numeric score for the relative merit of each of the options that satisfy all of the "givens"; assessing the risks associated with the preferred choices; and finally making a balanced judgement based on both the prioritisation and associated risks.

Problem Prevention:
A forward-looking methodology, (expanded in Failure Mode and effects Analysis) which is used to anticipate concerns and thereby develop either suitable preventative measures or actions to reduce the effects of the concern. It is used throughout the TOPS(8D) methodology to evaluate what could go wrong with an individual step, to assess what could be done to prevent the cause or to mitigate its effects and to identify who should take responsibility for necessary actions.

Figure C-II: Process Helps of Team Oriented Problem Solving

(III) Process Management

Process Management is a methodology for controlling and improving any process, although the emphasis in the programme is on manufacturing processes. Process Management is based on Deming's concept of 'Profound Knowledge' (Deming, 1993) where, for the continuous improvement of a system, two elements are required; an appreciation for the system, and knowledge of variation.

This concept is applied in Process management through four steps; a) identify and define the process; b) establish process management responsibilities; c) define and establish process controls; and d) improve process performance.

a) Identify and Define the Process: A process or system is defined as any activity that takes inputs and transforms them into outputs. Within the programme, emphasis is placed on the generality of process modelling, however the focus of the training is on the application of these concepts to manufacturing processes. Therefore inputs to a process are concerned with; materials; design; and resources, and the outputs are the resultant products. A process uses resources and operates within the constraints of controls. This step of process management establishes boundaries around the process by defining its inputs and outputs. In addition to the product, an important output of the process is the data that can be used to monitor and improve the process.

b) Establish Process Management Responsibilities: For any (manufacturing) process, there will be numerous people responsible for its operation. For effective process management however, a clearly defined process owner must be identified. The process owner will have an understanding of the process from beginning to end, the authority to effect changes, responsibility for process results, and a commitment to continuous improvement.

The process owner provides the focal point for the efforts of the team and is accountable for the performance of the process.

c) Define and Establish Process Controls: The fundamental principle of controlling a process is understanding the relationship between process variation which affect customer requirements and the factors of the process. Factors which significantly affect output variation are termed dominant sources of process variability (i.e. process set-up, time, incoming materials, and operators). Understanding of a processes dominant factors will determine the strategy for process management (i.e. a process may be controlled before the event (error proofing the process design), during the event (process controls) and / or after the event (defect detection).

Process control relies on collecting process data to understand its ongoing variability. By monitoring the variability and knowing when to react, the process can be controlled and improved.

This is achieved by using statistical process control techniques (SPC) derived from the work of Shewhart (1931). Shewhart control charts enable process operators to distinguish common (chance) causes of variation from and special (actual) causes of variation. This is important as the nature of the process variation will determine the type of quality improvement action to be employed; for special cause variation, problem solving efforts are required; and for common causes of variation process improvement techniques are required. Before common cause variation can be reduced, special causes must be eliminated to bring the process into a state of statistical control - where its performance is stable and predictable.

In addition to variations in process output, there will be variation due to the measurement instruments employed to collect process data. Consideration has to be given to obtaining accurate process data.

d) Improve Process Performance: The fourth step of Process Management is concerned with process improvement. To be of any consequence in terms of the quality of the product, this must be defined in terms of customer needs and expectations. Within the company, a seven stage Process Improvement methodology is advocated.

In the context of automotive manufacture, the finished product consists of over 500,000 components. Each has to be manufactured and assembled and each has characteristics which are important to the functioning of the vehicle. Controlling the processes by which the components are manufactured and assembled is essential to the overall quality of the product.

(IV) Failure Mode and Effects Analysis

Failure Mode and Effects Analysis (FMEA) is a method for anticipating potential problems, prioritising them, and preventing them or reducing the severity of their effects to reduce or prevent customer dissatisfaction.

The origins of FMEA lay in the aerospace industry (Samaio, 1995). Developed by NASA in the 1960's, FMEA was used to improve the reliability of aerospace, military, nuclear and electronic industry equipment and processes. Since the early 1970s, it has been used in the automotive industry as part of the product design and manufacturing planning disciplines, and it has been increasingly focused upon in the last 10 years by Western automotive manufacturers as a way of improving quality and reliability and reducing cost to compete with competition from Japanese manufacturers (Dale and Shaw, 1989).

The current methodology used in the automotive industry (SMMT 1989; SAE, 1994) is based largely on the American Military Standard MIL-STD-1629A (USA Department of Defence, 1980), although the company, as do other manufacturers (Aldridge and Dale, 1994), work to their own derivative of the methodology which forms part of the company's overall quality standard.

The FMEA methodology is applied during the product design or manufacturing planning stages of the engineering process. Concept or System FMEA studies and Design FMEA studies are employed to anticipate potential product design problems, and Process FMEA studies are employed to anticipate potential manufacturing process problems. FMEA is a time consuming task which, potentially, lends itself to automation (Price et al, 1992). At the time of the conception of the programme, FMEA remained to be a manual process.

The methodology comprises six stages; a) define scope and function; b) identify potential failure modes; c) prioritise potential failures; d) select and manage subsequent actions; e) observe and learn; and f) document the analysis.

a) Define Scope and Function: The FMEA methodology commences by defining the scope of the study, thereby determining the membership and structure of the study team. For Product designs, the functions of the product and defined, and in the case Process Failure Mode and Effects Analysis studies, the purposes of the manufacturing process are defined.

b) Identify Potential Failure Modes: Product or process failures are modelled in terms of causes, failure modes and effects, to assist in potential problem anticipation. The failure modes are categories of failure; function or purpose ceases; becomes intermittent; degrades or deteriorates; or becomes excessive. For each failure mode, effects of failure and potential causes of failure are anticipated.

c) Prioritise Potential Failures: A numerical rating is given for the probability of each cause occurring, the severity of each potential effect and the likelihood of the anticipated failure being detected during the product design phase or as part of the manufacturing process before reaching the customer. A product of the ratings for severity, cause and detection likelihood are used to establish a risk priority number, by which the anticipated potential failures, or problems, are prioritised.

d) Select and Manage Subsequent Actions: For the highest priority potential failures, actions are developed by the team which will either a) reduce the probability of failure, b) reduce the anticipated effects, or c) increase the likelihood of the failure being detected during the design phase or the manufacturing process. The actions developed by the team are implemented and the implementation is managed; monitored and appraised for effectiveness and progress.

e) Observe and Learn: After a product has been designed, or put into manufacture, the FMEA continues as a "living document" and serves as a forum to capture and consider actual failures of the product or process of which the lessons learned are incorporated into future product designs or manufacturing processes.

f) Document the Analysis: The Failure Mode and effects Analysis study, throughout its extended life (into the life of the product or manufacturing process) is recorded on formatted paper or using a dedicated software programme.

(V) Experimentation.

In the context of programme, experimentation provides for the engineer to gain knowledge about a particular product or process. According to Groves and Davies (1992), statistically designed experimentation is a methodology whereby many design changes can be made at once and conducting a series of tests and evaluations before decisions are made as to what next steps are taken in the development of the product or process.

A product (or process) can be modelled in terms of its function (or outputs) and the factors which affect it. By understanding the complex relationships which exist between factors and how they effect function (or output), engineers can exploit these relationships to improve product quality.

The experimental process is described in terms of the Deming cycle or Deming wheel (Scherkenbach, 1988) of plan-do-study-act. The 'plan' stage is concerned with selecting the function of the engineered system which is to be improved and identifying the factors which are likely to interact individually, or in combination, with the system function. To efficiently explore the factors (variables), an experimental plan (orthogonal array) is used which prescribes combinations of factor settings (levels). This determines the test plan. The 'do' stage is the implementation of the plan which involves running a series of test, with the 'study' stage being the analysis of the data generated by the tests.

Based on the analysis, the system factor levels are set to achieve the desired outcome ('act'). To confirm the findings, a confirmation test is conducted - which follows the PDSA cycle.

The same methodology can be applied to improve both product designs and manufacturing processes.

(VI) Quality Engineering

Quality Engineering is an embracing term which refers to an alternative approach to engineering. It represents a significant paradigm shift from the traditional approach to engineering. Conceived by Genichi Taguchi (Taguchi, 1986), the engineering philosophy is to ‘engineer in function, as opposed to engineering out problems’ (Groves and Davies, 1992). Taguchi advocates the integration of five key concepts into the engineering process; a) energy transfer; b) ideal function; c) signal to noise ratio; d) robustness; and e) quality loss function.

a) Energy Transfer: Taguchi offers a model of the engineered system whereby it its purpose is to convert inputs into outputs by the transfer of energy. Where all the input energy is converted into output, then no problems occur. If, however, not all input energy is converted into output, then the excess energy materialises as problems.

b) Ideal Function: Ideal function is where the system is functioning as intended; where all input energy is converted into output.

c) Signal to Noise Ratio: Taguchi distinguishes factors of an engineering system which are within the control of the engineer from those factors which are not. Factors which are beyond the control of the engineer, and which may include operating environment factors or system factors which are too expensive to control, are termed ‘noise factors’. The signal to noise ratio is an expression of a system’s output (signal) relative to its noise factors.

d) Robustness: The concept of robustness refers to a system’s ability to yield minimal functional variability in the presence of its noise factors (e.g. a car braking system operates in a variety of conditions and so one that is robust will be unaffected by its operating environment). This concept can be extended to manufacturing processes, where some process factors are regarded as noise factors.

e) Quality Loss Function: The quality loss function is a measure of the financial losses causes by an engineered system as it deviates from its desired output level. It takes account of the full life of a system and will include losses incurred long after the system was produced. The quality loss function encourages engineers to think differently about product specification; not seeing engineering as establishing tolerances, but focusing on achieving targets.

Quality engineering employs the disciplines of experimentation to explore the relationships between a systems control (signal) and noise factors. At the time the programme was conceived, Quality Engineering as defined by Taguchi was widely mis-understood (Brittle, 1990). The concepts were comparatively untested with the other methodologies included in the programme and as such Quality Engineering was a novel and challenging aspect of the programme.

(VII) Quality Function Deployment

Quality Function Deployment (QFD) (Akao, 1990) is a process for translating customer identified wants in a product into engineering specifications and subsequently into the product itself. Within the programme, QFD is presented within a wider framework termed Customer Focussed Engineering - the name which, perhaps, best describes the philosophy of this approach.

The definition and description of quality is explored further, with the introduction of Basic, Performance and Excitement Quality definitions. These distinctions between the various types of quality are necessary because of the underlying assumptions of the company's definition of quality; '"Products and services which meet the needs and expectations of customers".

The premise of QFD is asking the customer what he/she wants at the outset. However, recognising that the customer will not necessarily verbalise all of his/her wants and expectations, it has been recognised that there exists unspoken wants/expectations, hence the notions of basic and excitement quality. Basic quality refers to those characteristics in a product that the customer will expect (i.e. that the car will start; or the food will be non-toxic). Excitement quality refers to those characteristics which are yet undiscovered by the customer, but once realised will be the source of delight or excitement (i.e. a global positioning navigation system for car drivers -which, at the time of writing, was only something that would appear in a James Bond film).

The QFD approach is described in five phases; Phase 0 - Pre-planning (prioritisation of customer requirements); Phase I - Customer Requirements to Quality Characteristics; Phase II - Substitute Quality Characteristics to Design Characteristics; Phase III - Design Characteristics to Process Parameters; and Phase IV - Process Parameters to Production Requirements

Starting with phase 0, where customer requirements in a product are sought and prioritised, the QFD methodology progressively translates the customer requirements into the product design and the process by which the product is manufactured.

QFD provides an engineering framework for the deployment of the other quality methods considered in the programme.

(VIII) Team Building

The interpretation of team building was taken from a model developed by John Syer and Christopher Connolly (Syer and Connolly, 1987). Syer and Connolly's extensive background of working with sports teams to improve their performance through mental training influenced not only the terminology used to describe the skills, but the way in which they were trained. The team-building content of the people skills were eventually organised into six main topic areas; the effective team; roles and responsibilities; team process; the robust team; and right relationships

Set in the context of team meetings, the effective team considers how teams structure their time to achieve objectives and develop their team skills. Reflecting Connolly's and Syer's background within the sports world, the analogy of warming up and warming down is used to start and conclude team sessions. The process involves the attunement of individuals to the team and to the task, or objectives, which are to be achieved. Warming up is structured into six stages:- Place; Self; Others; Team; Purpose; and Activity (Table C-IV).

Stage	Process
Place	Place is concerned with the venue for the team meeting. This includes all aspects of the physical place (e.g. the size and layout of the room and the room temperature).
Self	Self is about the individuals which combine to form the team. It recognises that, as individuals, the team members arrive pre-occupied with previous events and issues. This is particularly relevant to engineers within the manufacturing industry where they can often go directly from one meeting to another, and although there in body, their minds are still at the previous meeting. The self activity of warm-up provides time for individuals to mentally put aside other issues (which are brought back at the end of the meeting as part of the warm-down). Exercises are facilitated during the training which include visualisation and the actual writing down of issues onto a piece of paper which is put away until the end of the meeting.
Others	Others is the first stage of getting to know the other team members. Often with meetings in a large company there will be people present who are not known to others at the meeting. It isn't uncommon to attend a meeting without ever finding out who some of the other people sat the meeting are. The others activity is about introductions to one or two people at the meeting. With mature teams, this activity is employed to allow members to get to know each other more, exploring more intimate aspects of the other people at the meeting as part of the building of team relationships.
Team	Team is concerned with the potential and identity of the team as a whole. The identity of a team is personified by the norms of behaviour within the team - the way team members conduct themselves, or behave, when they are together in the given team context. The team will have a common goal and a unique identity.
Purpose	Purpose is the task of the meeting - the reason the team have come together. At this stage in the warm-up, the process moves from maintenance of the team to task. In many meetings within the company, it is not uncommon for attendees to have different perceptions of the purpose of the meeting. This stage of the warm-up is to clarify, sometimes in a concise written sentence, the purpose of the meeting in terms of the intended outcome.
Activity	Activity is the final stage of the warm-up process. This is the 'how' to the purpose described above. Often it takes the form of an agenda, which gives detailed timing and a description of the process.

Table C-IV: Stages of Warming-up

Warming down is the reverse order of the 6 stages described; *activity* is reviewed, with outstanding items identified and actions to complete agreed; *purpose* is reviewed - did the team achieve the task of the meeting?; the maintenance stages of the warm-down are concerned with *team*, *others*, and *self* and are intended to review the team and personal aspects from the meeting; and *place* is considered - which normally means tidying the room before leaving it!

The duration of warming-up and warming down is determined by the length of the meeting. Clearly if the meeting is only 2 hours duration in total, then 10 minutes to warm-up and 10 minutes to warm-down would be appropriate. If the meeting is over several days, then more time can be given to warming-up and warming-down, with intermediate warming-up and warming-down processes at the start and finish of each day. Meeting duration is not the only consideration for determining the length of the warming-up and warming-down processes. For a new team, it may be necessary to spend more time on the maintenance stages of the process; developing the team identity and norms.

The concepts of warm-up are also advocated by Joiner (Scholtes, 1988) as a way for team members to transition from previous situations and issues into the team and the task of the meeting. Within the training design, this is presented as a checklist and taught as such, with the warm up and warm down activities systematically applied to the start and finish of each day of all training modules. Teams are also encouraged to employ this checklist as part of any group work which is undertaken during the training session.

A second aspect of the effective team is that of task; process; and maintenance. Within the organisation, increasing pressures on people's time and an overwhelming desire to 'get the job done' often result in team leaders focusing the meeting solely on task, with little or no regard to the maintenance and process of the meeting. Where attention is given beyond task, it is concerned only with process in a fairly limited way. Maintenance, in the sense of personal feelings and relationships, is not an aspect of team meetings which is considered, except within the context of a behavioural type training course.

Good task, process and maintenance facilitation requires a balance between the business part of the discussion and the team-building aspect. Team-building and the achievement of team goals should, ideally, happen simultaneously and should become an integral part of the team and team norms of behaviour when they're together as a whole, or as sub-teams.

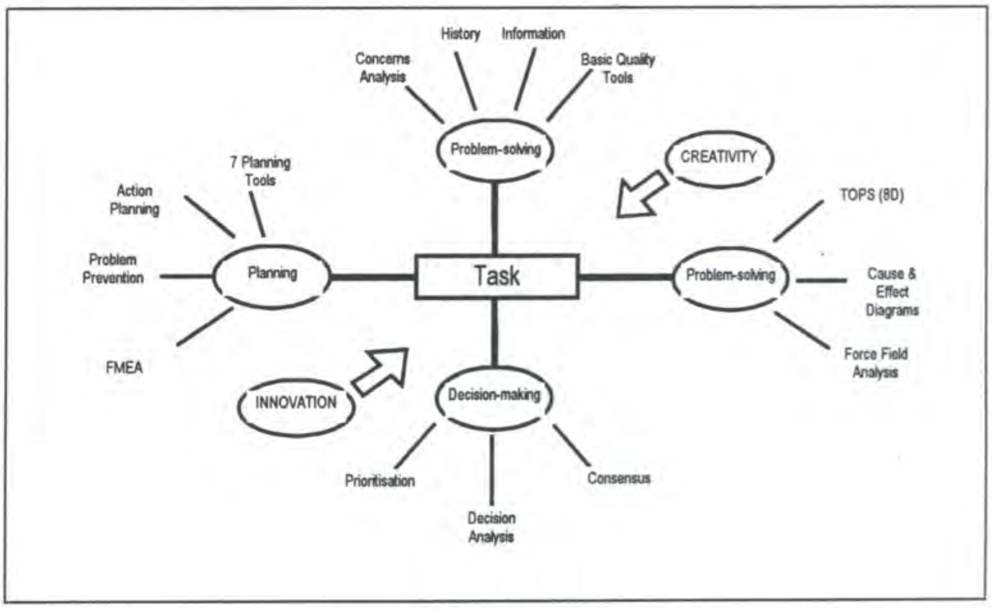

Figure C-III: Components of a task

Task, as the business part of a discussion, task will usually comprise the largest portion of time during the meeting. It will be described in the agenda for the meeting, although the agenda, if structured properly, will reflect good process design. Within the training, task is described in terms of its components using the illustration (mindmap) in Figure C-III.

The maintenance part of the meeting (see Figure C-IV) is concerned with the feelings of team members and the warm up and warm down activities of the meeting are concerned with maintenance.

The process, the way in which the meeting is conducted, provides the framework for the activities of the meeting.

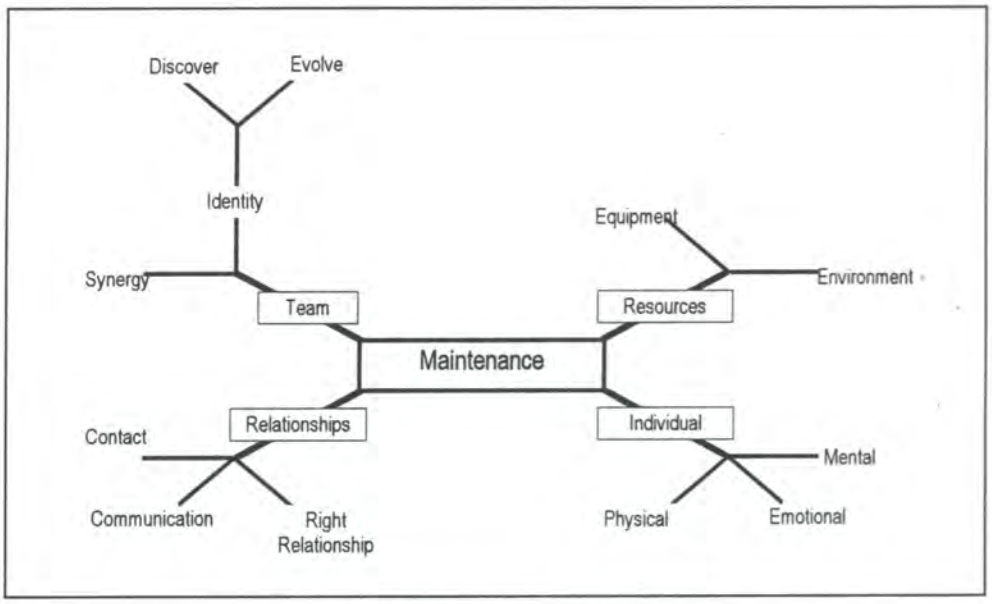

Figure C-IV: Influential factors in maintenance

The second area of team building is that of roles and responsibilities. The assemblance of a group of engineers brought together to address an engineering problem, to plan a project, or simply to share ideas and learn, can be effective or ineffective depending largely on how the meeting is run and how the team works together.

Within organisations, membership of a team is largely based on technical knowledge and skills required to achieve the desired result (Larson and LaFasto, 1989) and little consideration is given in the selection process to the team dynamics requirements. The concepts of task, process, and maintenance are advocated to provide a framework within which the team operate and the roles played by individuals are also considered to be important to the success of the team.

Within any group, individuals are expected to follow identified behaviour patterns (Briggs-Myers & Myers, 1993) which , ideally, align to the roles required for effective team operation (Belbin, 1993). The team roles taught within the programme are described in Table C-V.

Team Role	Role Description
Team Leader	The person responsible for leading the team and encouraging them to the achievement of the task through setting the team's aims and objectives. The leader also acts as the team's spokesperson; reflecting and representing the views of the team. The team leader role should remain stable for the duration of the team's task.
Facilitator	The primary function of the facilitator is team building. The facilitator is responsible for the team's maintenance and observing team performance. The team facilitator can be any team member and the role may be played by different members from meeting to meeting
Time Manager	Responsible for planning and managing timing changes, the time manager role can be played by any team member. The time manager will often facilitate timing changes to an agenda and remind the team of timing issues affecting the meeting.
Scribe	The scribe, or recorder, serves the team by recording meeting events of significance, such as ideas and decisions. The role of scribe can easily be abused as the person playing the role has the power to interpret ideas and decisions when recording them. Also, where a discussion requires the continuous noting down of comments, the scribe will often find him/herself standing at the front of the room and acting as the focal point of discussion. Emphasis is placed on recording what is actually said, as opposed to recording the scribe's interpretation of it.

Table C-V: Team Roles

Attitudes are considered briefly within the programme as they affect teamwork by shaping how a person perceives another person or a situation. They provide a mental frame of reference, based on personal experience, with which to interpret events and guide the sequence response to them.

To raise awareness of the effect of attitudes and to serve as a facilitation tool, a simple 'see, imagine, feel' technique is included in the programme. The technique is intended to make explicit perceptions of a situation and the response it evokes. It slows down the process of a meeting, particularly where feelings among team members are negative, with the objective of preventing inappropriate actions and reactions to situations.

Termed as a 'process-check' in the training material, team members are encouraged to; state their observations (what they see); describe how they interpret their observations (what they imagine); and describe how this makes them feel. By making perceptions explicit, situations can be discussed and understood.

The third and fourth topic areas of team building (team process and robust team) draw on concepts from aspects of the technical content of the programme. Team process applies principles of process management to the management of team behaviour. The four step process management model is used as a framework for using team-building and communication skills, with emphasis placed on assigning team roles and responsibilities and monitoring team behaviour through descriptive feedback. The concepts of common and special causes of variation are used to distinguish types of behaviour within a team.

Similarly to team process, the concept of robust team uses principles and models of quality engineering as a framework for using team-building and communication skills to build and maintain and effective team. Team process and robust team are taught as part of their respective technical modules.

An important factor of effective team work is trust (Syer, 1986) and 'right relationships' is an eight stage team-building process to develop trust between team members. The eight stage process (awareness of self; appreciation of self; awareness of others; appreciation of differences; contact; communication; respect; and trust) employs the speaking, listening and framing communication skills and uses descriptive feedback and questioning to facilitate the team through the eight stages. These are described in the next section.

(IX) Communication

The second major topic area of the People skills; communication, is structured into 6 sections; listening, questioning, descriptive feedback, speaking guidelines, and framing information.

The emphasis is on verbal communication and is primarily concerned with the transfer of information and the common interpretation of that information. Each section encapsulates the concepts into techniques which are useable by the participants of the program.

Listening, or active listening as it is more commonly known, is included in the program to address two main concerns identified in the communication which takes place at business meetings. The first is that listeners fail to understand the message that has been relayed to them through verbal communication. The second is that listeners ignore the message offering an alternative idea or statement.

Listening is presented using a simple model (listening circle) which comprises of two elements or sub-processes; restating a message to check understanding; and building on ideas, or statements. The listening cycle is modelled within the training as '3 likes and 1 wish'; having listened to the idea, the listener restates and gives 3 things he / she liked about the idea. This has the desired effect of; encouraging the listener to listen and understand the idea; to think about its merits; and to avoid the temptation of immediately criticising or offering a counter-idea. This approach is also intended to effect the mutual ownership of ideas.

After the statement of the things that are liked about the idea, comes the suggested improvement. These are expressed as 'wishes' and are intended to build on the original idea or statement. Again this is intended to open a discussion which is constructive, as opposed to the destructive discussion of ideas which is often seen in business meetings.

The second section of communication is questioning which is drawn largely from Hargie (1988) and is divided in two main areas; questioning for content; and questioning for understanding. Questioning for content considers the basic principles of vocabulary and expression used and the way in which questions are structured, recognising that the way in which a question is asked will, to a large extent, determine the answer which is given. Good questioning practice requires that vocabulary is; neutral; minimises assumptions and implicit judgements; is not too complex conceptually; and has simple meaning with clear intentions.

With regard to structure, closed and open questions are considered. Closed questions are those which can only have a limited range of answers (i.e. yes/no; alternative; multiple choice). The answers which can be elicited using closed questions are limited and will be framed completely in the language of the questioner. The use of closed questions is advocated in the programme to clarify and pin-point information.

Open questions however, require the respondent to be more descriptive and can be narrative or directive. Open-narrative questions ask the respondent for a description or account as part of their answer and provide the opportunity for an unrestricted flow of information, drawing on the knowledge, experience and opinion of the respondent. Open-narrative questions provide a rich source of data. Open-directive questions are more specific as they focus on more precise information and knowledge.

Whilst relatively simple, questioning for content can be a powerful tool in an engineering environment For example, in a problem solving meeting where the nature of the problem is often technical and complex, it is easy for team members to use complex vocabulary in the formulation of problem solving questions. By emphasising these basic principles, the program is attempting to encourage engineers to simplify their thinking by simplifying the questions which they ask.

Questioning for understanding builds on from content-questioning in that it explores the effect of an individual's experience, values and judgements, or the context, of the answer given to a question or as a statement. Where, for example, engineers are defining customer wants, each customer surveyed will have their own personal experience and opinion of what constitutes good vehicle attributes. Although these will be reflected in their answers, they are not necessarily made explicit and it is therefore necessary to explore further using understanding-questions.

The third area of communication; descriptive feedback refers to giving and receiving information about the performance of an individual or a group. The emphasis is on description, as opposed to judgement, as a method of giving feedback. By stating observations of behaviour, the person receiving feedback is able to understand what his/her behaviour was, how it affected others, and give them something tangible to act upon. The receiver of feedback is also less likely to disagree with observations.

For the person giving feedback, being required to describe behaviour encourages observation and places the giving of feedback on an objective plain. When giving feedback to a close work colleague, the objectivity of the feedback makes the interaction easier for both parties (Cairns, 1989).

The fourth area of communication; speaking guidelines is intended to facilitate the communication process between team members. Based on Hargie (1991) and intended to encourage engineers to take ownership of ideas and criticism, a simple framework of nine guidelines is provided; talk from personal experience, speak to not about people present, address the person by name, look at the person you're talking to, say 'I', not 'we', make statements before questions, trace opinions back to observations, describe don't judge, accept that feelings will contribute to the discussion, and say 'I would', rather than 'you should'.

In conjunction with speaking guidelines, the concept of framing information is advocated to provide structure for information to be communicated (Table C-VI). This becomes particularly important where complex engineering discussions are taking place between team members, not all of whom will be familiar with the subject.

Frame	Guideline
Headlines	Up-front statements which capture the main topic of the information to be communicated and capture the attention and set the context of the information to be transmitted by the receiver. The use of headlines in a newspaper are an illustration of this guideline.
Signposts	Similar to headlines, but feature early in the communication and are intended to provide the receiver of information with an indication of the topics and their order which is contained in the information to be communicated. As the name suggests, signposts alert the receiver as to where the communication is going.
Boundaries	The breaks where one topic ends and a new one begins. They provide markers for the receiver of information to stay in touch with the communication.
Key points	Provide a concise summary in a headline format of the information that has been communicated. Key points come towards the end of a communication.
Links	For complex information, it may be necessary to make explicit any linkages between topics which have been communicated and/or to topics which have previously been addressed.

Table C-VI: Framing Information

(X) Implementation

The third main topic of people skills is implementation and is concerned primarily with bringing about change. The topic is structured into; force field analysis; action planning; decision-making; and change agency.

Force field analysis is set in the context of systems thinking (Senge, 1990) with the emphasis on viewing the overall process as opposed to the actual content. Developed by Kurt Lewin in the 1930's from his field theory (Checkland, 1981), force field analysis models phenomena as a system with forces acting upon it. Whilst all the forces are in equilibrium, the system remains stable and does not change, but when one force increases or another decreases, then change in the system will result.

Force field analysis provides engineers with a technique for considering the wider picture and identifying opportunities for implementing change. When planning a change to a process (i.e. the introduction of SPC to a process), the current state can be modelled in terms of its driving (i.e. include the desire of the process operators to improve quality, reduce scrap levels, avoid re-work) and restraining forces (i.e. lack of knowledge of SPC, its poor track record in other processes, the extra time and effort it would take to introduce) which are keeping it in a state of equilibrium. Driving forces are those which, if increased would move the system towards the desired or intended state. Restraining forces are those which are keeping the system from moving to the desired state.

Action planning provides a methodology for achieving specific objectives through the implementation of clear decisions. Action planning is undertaken within a clear agenda where substantial issues require follow-up. For the purposes of the programme, the following set of rules were devised.

1. *The issue must have an owner* - the owner will usually be the person who knows most about the concern and will be responsible (and accountable) for implementing the plan and any necessary follow-up actions are taken.

2. *A clear time frame is set for the action planning*- action planning sessions can take from a few minutes to several days. It is important however that a predetermined time frame is established and adhered to. By their creative nature, action planning sessions if left unchecked can overrun team members time expectations and lead to frustration.

3. *A small team should be used* - 3 - 5 team members recommended. Where a team is too large, there is a risk that not all team members will be fully involved.

4. *Someone other than the owner of the issue should facilitate the process* - as the owner is the person who possesses greatest knowledge about the issue or concern, he/she should be fully engaged and not thinking about process issues. Further, the owner needs to remain open to the ideas which emerge, leaving facilitation issues to someone else.

5. *Quality matters more than quantity* - action planning is a refining process which cultivates new ideas. Action steps should be few, clear and practical.

6. *Participants contribute ideas for the benefit of the owner of the issue* - team members should recognise that the owner of the issue decides which strategies are best and should avoid becoming attached to their ideas to the detriment of the action plan.

7. *Creative ways of generating ideas* - action plans are only as effective as the quality of the ideas within them. Judgement of suggestions should be suspended to enable team members to take unusual mental journeys in the teams quest of action strategies.

Action planning integrates the communication skills of; listening and restatement; framing information; and speaking guidelines as part of a three stage process of generating, refining, and implementing plans. The process facilitates the generation of ideas from team members which are refined in to a plan and implemented within set time constraints.

The third area of the implementation topic recognises the difficulty of team decision making, where a range of views are held by team members. Five types of decision making processes are considered; a) unilateral; b) polling; c) prioritising; d) compromise; and e) consensus. These are outlined in Table C-VII.

Decision Type	Outline Description
a) Unilateral	The simplest form of decision making, unilateral decisions are those taken by an individual or team leader. The process has the advantage of cutting through confusion and is quick. These types of decisions can be counterproductive however, as people affected by the decision have little or no involvement in the process. In team situations, unilateral decisions should be avoided as they undermine the team's identity.
b) Polling	Polling is a voting process to reach a decision. It has the advantage of enfranchising participants and is democratic. The main disadvantage with this process is that a small majority will 'win' the vote and the decision may disempower team members.
c) Prioritising	Prioritisation is a rational decision making process as it is based on assessment against chosen criteria. It takes time and has the advantage of allowing decision makers to undertake a careful analysis, identifying criteria internal as well as external to the team. In some situations, the time taken to come to a decision may be a disadvantage and it can become impractical where large decision making groups are concerned.
d) Compromise	Compromise decision making takes account of the positions of all parties involved and affected by the decision making process. Its main advantage is overcoming an impasse in a team's progress, however it tends towards simple decisions and it is difficult to mobilise a team to actions reached through compromise.
e) Consensus	Consensus decision-making is a development of compromise decision-making. The interests, and not positions, of those involved in the decision making process are considered, with the aim of finding solutions which meet the needs of all concerned. Consensus decision making is suited to major policy decisions which have long term effects. Consensus decision making can be very time consuming and requires decision makers to put aside individual differences for the sake of the team, which makes the process is vunerable to sabotage.

Table C-VII: Decision Making Processes

The principle aim of the training programme is to bring about change by providing engineers with different ways of designing and developing products and processes to improve quality. In this sense, the programme is subject to factors which influence the diffusion of new ideas in any culture. Based on Rogers (1983), change agency identifies people affected by change as innovators, early adopters, early majority, late majority, and laggards. Innovators create change and laggards are reluctant to follow change. The factors which affect the adoption of change are categorised into; perceived attributes of innovations, type of innovation-decision; communication channels; nature of the social system; and scope of change agents efforts.

The programmes trainer-consultants and participants are all potential change agents. By understanding the process by which innovations spread; knowledge; persuasion; decision; implementation; and confirmation, they are encouraged to play a positive change agent role in the adoption of the programme content within the engineering community.

As participants in the programme, engineers are expected to return to the workplace and introduce new practices to an environment where they are likely to face some resistance. Advocating change provides a four step negotiation process designed to reach an agreement when two parties have some interests which are shared and others that are opposed. The four stages are 1) Giving people equal time to the problem; 2) Focus on interests and not positions; 3) Find new solutions so that everybody wins; and 4) Insist on using objective criteria.

The emphasis of the process is exploring the unstated interests people may have which motivate them to adopt a particular stand, or position, with respect to changes. By explaining why they have taken a particular stand, common and opposing interests are revealed allowing effort to be focused on the areas which are different.

(XI) Innovation

The fourth topic area of the people skills is concerned with innovation methodologies of; thinking; scientific methodology; creative thinking strategies; idea Mapping; cognitive mapping; language mapping; conceptual block-busting; brainstorming; paradigm Shifts; and innovative product development.

The topic area is intended to challenge the habitual ways in which the company's engineers think. Ornstein (1972) identifies two main modes of thinking which can be generally described as analytical and associative. Analytical thinking is rational and linear (left brain thinking) and associative thinking is more visual, imaginative and non-rational. (right brain thinking)..

De Bono (1991) identifies divergent and convergent types of thinking. Divergent thinking is moving away from responses already known to generate a number of answers to a problem. Brainstorming is an example of divergent thinking. Convergent thinking is traditional logical thinking which strives to proceed directly from one state of information to another.

Both the models of thinking are explored in the programme and intended to provide fundamental building blocks as part of a conceptual foundation for this topic area.

Scientific methodology is a high level examination of the disciplines of scientific discovery. The notions of logical and deductive reasoning as rational processes of knowledge development from theory and universal laws of science are reviewed. The concept of paradigms are recognised and their implications to innovative product and process development are explored.

The inclusion of scientific methodology in the programme is intended to provide engineers with an insight into scientific discovery and relate lessons from this field of activity to their own work. Again, this section of the innovation topic area of the programme, along with thinking and creative thinking strategies, provides further building blocks of a conceptual foundation for the tools and techniques which are explored in the remainder of the innovation section of the people skills.

The notions of experimental and exploratory thinking are presented as contrasting thinking styles. Experimental thinking involves the search for a right answer. It calls for the testing of specific hypotheses and employs ready-made formulae and statistical analysis to manipulate data to find a solution to a problem. This type of thinking is linear in nature and is a process of orderly steps.

Exploratory thinking refers to the search for new ideas and involves associative, as opposed to analytical, thought processes. Exploratory thinking attempts to expand the options available as solutions to a problem in the quest of finding innovative solutions which are better than traditional ones.

Jumping out is a selection of techniques intended to make the thinker (engineer) create new associations by cutting across established patterns of thought. The techniques are initiated when the group reach an impasse in their activity and this is often characterised by feelings of frustration among team members. Jumping out is based on the principle of discontinuity. Discontinuity is switching from one framework of knowledge or expertise into another. It is the process of discontinuing one's thinking in a particular frame of reference and thinking in terms of an alternative frame. Discontinuity is accredited, among other things, with the creation of Chaos theory; where James Yorke, a physicist, and Peter May, a mathematician, made connections between their respective fields which led to the conception of chaos theory.

Once an impasse situation has been recognised, three jump-out options are available to the team; 1) random stimulus; 2) what-if; and 3) advanced visual thinking.

Random-stimulus technique applies the principle of discontinuity by introducing a word or picture which is random or irrelevant to the discussion or problem. Within a specified time-frame, associations between the random word (picture) and the problem are discussed, without criticism of suggestions and generating as many suggestions as possible. Ideas are recorded onto a flip-chart. Once the time limit is up, suggestions are revisited and considered as to whether they can be developed into a solution.

The what-if technique is based on Koestler's model of bisociation (Koestler, 1964) and applies the principle of discontinuity to the creative process by considering an opposite or outrageous solution to a problem. An example might be square wheels to improve vehicle ride and handling. In order to make this solution work, the team would have to engage in thinking about square wheel countermeasures - an activity which may spur creativity to solve the original problem.

The final set of techniques considered in the programme which use discontinuity as a creative device, are advanced visual thinking skills. Addressing the thinking error of misplaced concreteness; where something is believed to be true beyond any doubt, when in fact it is not, advanced visual thinking allows the creator to step out of these strongly held beliefs. The concept is accredited (Arnheim, 1969) with Einstein's theory of relativity; where Einstein imagined he was riding a beam of light like a surfer on a wave, and the discovery of the molecular structure of benzene; where the German chemist Kekule dreamt of snakes eating their own tails.

Three techniques are offered; Operator STC, which applies the principles of exaggeration distortion; the Model of Miniscule Dwarfs, which involves humanising a machine as if each component were a small dwarf to think about its interactions; and Ideal Final Result, which involves jumping to the ideal solution and disregarding constraints.

As a method for organising information, idea mapping enables individuals or teams to relatively quickly assemble information in a logical structure which reflects the relationship and associations between the pieces of information. Based on the work of Buzan (1993), idea mapping represents information and concepts without using long pieces of text.

Within an engineering context, where often highly complex issues are addressed, idea maps can be used to represent and communicate the whole picture and although the detail that can be achieved through the use of passages of text is lost, the advantages of speed, particularly in team situations, outweigh the disadvantage of loss of detail.

The basis for cognitive mapping lay in George Kelly's (Kelly, 1955) personal construct theory which proposes that human behaviour is determined by individual theories about the world (personal constructs) that are used to make sense of the world and anticipate what will happen in a given situation. These individual theories are the rejected or retained on the basis of how accurately they predict events.

Cognitive mapping uses several techniques (repertory grids, visual card sort technique) to represent and communicate individual's perceptions of phenomena. In an engineering context, this may the customers perception of the value or quality of a product, or the perceptions of problem solving team members of a problem.

Cognitive mapping can be used to collect customer information or to share the ideas (perceptions) of team members.

As an extension to paradigms and personal constructs topics of the programme, conceptual block-busting is concerned with overcoming these as barriers to innovation. The pre-conceived ways in which individuals (engineers) perceive the world are said to act as conceptual blocks or barriers to both the correct situation at hand and the appropriate response to that situation.

Conceptual block-busting is intended to provide the engineer / engineering teams with the skills to identify these barriers to innovation and then to overcome them. To aid understanding, the conceptual blocks are categorised into; perceptual, attitudinal, emotional, and intellectual blocks and mental strategies offered to overcome these blocks.

The term brainstorming is accredited to Alex Osborn and refers to a creative process where a prime inhibitor to creativity - judgement is suspended. The purpose of brainstorming is to generate as many ideas as possible before embarking on a process of appraisal and refinement. To stimulate ideas, an appropriate question is used.

To facilitate the process, a set of rules were developed based on the experiences of the programme development team. These were intended to facilitate effective brainstorming by a team of engineers. Apart from the suspension of judgement, a rule was included for establishing and adhering to a time limit. This time limit set a maximum and minimum time for the team. The time limit was not to be exceeded, but equally important it was to be achieved. Although this could, and often did, result in long periods of silence at the end of a brainstorming session, it also encouraged people to think deep when their ideas had dried up.

As part of the brainstorming session, ideas were recorded as stated - often using a metaplan board and cards, which aided the appraisal and refinement process.

Paradigms (Kuhn, 1962) form the basis for a shared perception of the world, providing a coherent structure which enables individuals to share information. Paradigms are also counterproductive as when people fail to shift from one paradigm to another (cognitive dissonance), they are left in framework which does not recognise new data for what it really is - or as it is determined in the new (better) paradigm.

To shift paradigms is to recognise their existence. Within industry, paradigms affect the perceptions of customers and of engineers. From recognising and understanding paradigms, strategies. Such as conceptual block-busting can be employed to overcome them as inhibiting factors to innovation.

Within the programme, the innovation concepts integrated to form Innovative Product Development; a three stage product development process of; 1) product definition, 2) product generation, 3) product refinement and discrimination. At each stage of the process, the engineer is offered a selection of the techniques to employ to facilitate product development.

Appendix D Pilot Review Questionnaire

Section I

The programme has included behavioural skills (Team-building etc.) and Technical skills (quality methods and techniques). Both are considered to play an essential part in quality improvement.

The following questions address content issues / opportunities.

In your opinion rate (place an 'X' along the line) the overall balance between:

1. Behavioural and Technical skills?

Behavioural Skills										Technical Skills
5	4	3	2	1	0	1	2	3	4	5
Too Much					Just Right					Too Much

2. Lecture and syndicate work?

Lecture										Syndicate Work
5	4	3	2	1	0	1	2	3	4	5
Too Much					Just Right					Too Much

3. Conceptual knowledge and application training?

Knowledge										Training
5	4	3	2	1	0	1	2	3	4	5
Too Much					Just Right					Too Much

Section II

How much will the following aspects of the programme contribute to your professional development and ability to improve the quality of your company's products?

Answer in terms of both short and long term contributions.

	Short Term	Long Term
	Not Important — Very Important	Not Important — Very Important
1. Behavioural skills?	1 2 3 4	1 2 3 4
2. Technical Quality skills?	1 2 3 4	1 2 3 4
3. Conceptual Knowledge?	1 2 3 4	1 2 3 4
4. Application Knowledge?	1 2 3 4	1 2 3 4
5. Training Materials?	1 2 3 4	1 2 3 4
6. Case Study?	1 2 3 4	1 2 3 4

Section III

Please rate the programme modules for content, relevance, delivery and training materials, etc.

Module 1 Foundation

	Level I	Level II
	Bad Good	Bad Good
Content	1 2 3 4	1 2 3 4
	Bad Good	Bad Good
Relevance	1 2 3 4	1 2 3 4
	Bad Good	Bad Good
Delivery	1 2 3 4	1 2 3 4
	Bad Good	Bad Good
Training Materials	1 2 3 4	1 2 3 4
	Bad Good	Bad Good
Case Study	1 2 3 4	1 2 3 4
	Bad Good	Bad Good
Application to workplace	1 2 3 4	1 2 3 4

Module 2 TOPS (8D)

	Level I	Level II
Content	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Relevance	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Delivery	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Training Materials	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Case Study	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Application to workplace	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good

Module 3 Process Management

	Level I	Level II
Content	Bad — Good 1 2 3 4	Bad — Good 1 2 3 4
Relevance	Bad — Good 1 2 3 4	Bad — Good 1 2 3 4
Delivery	Bad — Good 1 2 3 4	Bad — Good 1 2 3 4
Training Materials	Bad — Good 1 2 3 4	Bad — Good 1 2 3 4
Case Study	Bad — Good 1 2 3 4	Bad — Good 1 2 3 4
Application to workplace	Bad — Good 1 2 3 4	Bad — Good 1 2 3 4

Module 4 FMEA

	Level I	Level II
Content	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Relevance	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Delivery	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Training Materials	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Case Study	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Application to workplace	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good

Module 5 Experimentation

	Level I	Level II
Content	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Relevance	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Delivery	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Training Materials	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Case Study	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Application to workplace	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good

Module 6 Quality Engineering

	Level I	Level II
Content	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Relevance	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Delivery	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Training Materials	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Case Study	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good
Application to workplace	Bad 1 2 3 4 Good	Bad 1 2 3 4 Good

Module 7 Customer Focussed Engineering

	Level I	Level II
Content	Bad Good 1 2 3 4	Bad Good 1 2 3 4
Relevance	Bad Good 1 2 3 4	Bad Good 1 2 3 4
Delivery	Bad Good 1 2 3 4	Bad Good 1 2 3 4
Training Materials	Bad Good 1 2 3 4	Bad Good 1 2 3 4
Case Study	Bad Good 1 2 3 4	Bad Good 1 2 3 4
Application to workplace	Bad Good 1 2 3 4	Bad Good 1 2 3 4

Section IV Overall Assessment

1. Based on your knowledge of the activities in your area (Design, Manufacturing, assembly, support), identify any behavioural or quality skills that have not been included in the training you received.

____

____

____

____

2. Based on your knowledge of the activities in your area (Design, Manufacturing, assembly, support), identify any behavioural or quality skills that should be eliminated or receive less emphasis in future training.

____

____

____

____

3. Identify three aspects of the programme that improved your skills.

____

____

____

4. Identify three areas of the programme that require further improvement.

____

____

____

Which area do you work in?
Product design, Manufacturing Engineering, Program Office, etc.

____

Which parts of the programme do you believe it is important for your colleagues from your area to attend?

All Level I	All Level I Some Level II	All Level I All Level II

Appendix E Knowledge Gain Questionnaires

Engineering Quality Improvement Program Foundation Module

Pre-Questionnaire (Alpha form)

This questionnaire is anonymous. However, because we need to relate the "before" and "after" experience, we would appreciate you coding the "before" and "after" answer forms with your programme delegate number.

Please attempt to answer the questions overleaf. Record your answers on the enclosed (2-sheet) answer form, following the instructions given below. When you have completed the assignment, tear off the top sheet of the form, fold it and post in the 'voting box' provided. In order to ensure anonymity do not write your name on the answer form (see above). The result of this assessment will remain confidential to this group. You should spend approximately 20 minutes on this assignment. Please do not consult with, or talk to, your colleagues whilst completing the assignment.

Instructions for Answering Questions

(a) There are 5 answers marked A to E in the key beneath each question. **Each question has only one correct answer.** Indicate your answer by pencilling a horizontal line across the appropriate cell in the row corresponding to the question number being answered.

For example,
if you believe the correct answer to be option D, then your answer form should be marked as:
Q1 (A) (B) (C) ~~(D)~~ (E) (?)

(b) If you do not know the answer the answer to a question, pencil across the 'don't know cell "(?)"

For example,
if you do not know the answer to question 3 then your answer form should be marked as:
Q3 (A) (B) (C) (D) (E) ~~(?)~~

In order that this assessment is a true reflection of your performance, you are requested not to guess the answer to any question where you do not know the answer.

(c) If you wish to change any of your answers, cross out any cells inadvertently pencilled across.

For example,
if you wish to change your selected option for Q1 from (D) to (B) your answer form should be marked as:

Q1 (A) ~~(B)~~ (C) ~~(D)~~ (E) (?)

1. Cause and Effect relationships can be shown on a:

Key A. Ishikawa diagram
B. Pareto chart
C. Stem and leaf plot
D. Flow chart
E. control chart

2. In Phase II of a 4 phase QFD, Substitute Quality Characteristics (SQC's) are translated into:

Key A. Customer wants
B. Rates of importance
C. Design characteristics
D. Process Parameters
E. Production controls

3. The performance of Dagenham and Valencia plants are compared by considering the value of the Process Capability index Cpk for a specific product characteristic. The value of Dagenham is 1.6 and that for Valencia 2.0. Which process produces the higher quality?

Key A. Dagenham
B. Valencia
C. Neither
D. Both equal
E. Can't say without more information

4. A confirmation run should always be used for:

Key A. Control charting
B. FMEA
C. QFD
D. Designed experimentation
E. Pareto analysis

5. As part of the production planning phase of QFD which of the following techniques is <u>not</u> appropriate?

Key A. xbar/R Chart
B. Gauge capability
C. Action planning
D. Process flow diagram
E. Force Field analysis

6. Listening skills are demonstrated by:

Key A. Saying "I've understood"
B. Maintaining eye contact
C. Using restatement
D. Nodding one's head
E. Not adding your own ideas

7. In the Kano Model of Quality Features, which of the following will result in customer dissatisfaction if not fully achieved?

Key A. Unspoken quality
B. Basic quality
C. Performance quality
D. Excitement quality
E. One-dimensional quality

8. EAO, Mazda and NAAO Have tolerances of $\pm$ 0.25mm, $\pm$ 0.70mm and $\pm$ 0.50mm respectively on an identical stamping characteristic. Who is likely to produce the lowest quality in the longer term, measured in terms of this characteristic:

Key A. EAO
B. Mazda
C. NAAO
D. All equal
E. Can't say without more information

9. A design verification plan is used by a Design FMEA team to:

Key
- A. calculate the risk priority number (RPN)
- B. assess the severity rating
- C. assess the occurrence rating
- D. assess the detection rating
- E. summarise product functions

10. A Pareto diagram can be used to:

Key
- A. Show frequency distributions of variable data
- B. Prioritise action
- C. Assess correlation
- D. Control a process
- E. Identify process flow

11. All eight disciplines of the TOPS (8D) problem solving process should be considered:

Key
- A. On all concerns discussed at the Quality Strategy Committee
- B. When the resolution of a problem is beyond the capability of an individual
- C. Where a problem has recurred
- D. On every concern with more than 5TGW/100
- E. When the cause of concern is known

12. The task element of a meeting does not consider:

Key
- A. Team process
- B. Problem prevention
- C. Project planning analysis
- D. Cause and effect analysis
- E. Prioritisation

13. Addressing negative quality is the primary concern of:

Key
- A. QFD
- B. FMEA
- C. Customer survey
- D. Quality Engineering
- E. Increasing customer satisfaction

14. One of the major elements of TQE is:

Key
- A. ISO 9000
- B. Conformance to specification
- C. People
- D. 100% inspection
- E. Q101

15. In performing a 'Taguchi' style experiment which of the following is <u>NOT</u> of prime concern?

Key
- A. Changing factor level
- B. Minimise the use of energy
- C. Identifying control factors
- D. Minimise the effect of noise
- E. Eliminating an error state

16 Time spent paying attention to the feelings and relationships of team members is:

Key
- A. Maintenance
- B. Process
- C. Group dynamics
- D. The team leader's responsibility
- E. Should be an agenda item

17. Effective process ownership is dependent on the process owner having:

Key
- A. Knowledge of Experimentation
- B. Intimate process knowledge
- C. Management Roll (MR) status
- D. Knowledge of SPC
- E. Responsibility for maintaining control charts

18. Which of the following would you expect to see as part of the D2 - describe the problem discipline in TOPS(8D):

Key A. Risk assessment worksheet
B. Root cause identification
C. Is / Is Not information
D. Interim Containment Action (ICA)
E. Concern analysis

19. During the Warm Down process at a meeting, the team should:

Key A. Establish an action plan
B. Prepare the agenda for the next meeting
C. Prepare the minutes of the meeting
D. Deal with unfinished business
E. Assign tasks to be completed before the next meeting

20. Which of the following techniques are appropriate to apply in an effort to reduce common cause variation in a anufacturing process?

Key A. Customer Focussed Engineering
B. Experimentation
C. SPC
D. FMEA
E. TOPS(8D)

21. Choose the statement which best demonstrates the advantages of using descriptive rather than evaluating feedback:

Key A. It lets people know when they have gone wrong or made a mistake
B. People are given information about their behaviour without judgement and without evoking defensive responses
C. It avoids having to commit oneself to a position of judgement about others, thereby avoiding conflict
D. Evaluating feedback doesn't give the individual the option of changing behaviour
E. Descriptive feedback avoids discussion and wasting time because responses from the other person are avoided

22. During the design deployment phase of QFD, which of the following is appropriate?

Key A. Defect detection
B. Customer survey
C. Design FMEA
D. Process FMEA
E. xbar/R chart

23. In Taguchi's Quality Engineering approach, the Parameter Design stage is concerned with:

Key A. Upgrading product or process parameters
B. Maximising energy transfer
C. Selecting technology for function
D. Selecting nominals for robustness
E. Minimising the loss function

24. In a team meeting, attention to team members' emotional and physical needs is the main responsibility of:

Key A. The champion
B. The team leader
C. The facilitator
D. The scribe
E. The time manager

25. A process is not in control when:

Key A. Only common cause variation is present
B. It is outside of specification
C. When special causes of variation are present
D. It only has points within the control limits
E. Cpk <1.33

Engineering Quality Improvement Program
Foundation Module

Post-Questionnaire (Beta form)

This questionnaire is anonymous. However, because we need to relate the "before" and "after" experience, we would appreciate you coding the "before" and "after" answer forms with your programme delegate number.

Please attempt to answer the questions overleaf. Record your answers on the enclosed (2-sheet) answer form, following the instructions given below. When you have completed the assignment, tear off the top sheet of the form, fold it and post in the 'voting box' provided. In order to ensure anonymity do not write your name on the answer form (see above). The result of this assessment will remain confidential to this group. You should spend approximately 20 minutes on this assignment. Please do not consult with, or talk to, your colleagues whilst completing the assignment.

Instructions for Answering Questions

(a) There are 5 answers marked A to E in the key beneath each question. **Each question has only one correct answer.** Indicate your answer by pencilling a horizontal line across the appropriate cell in the row corresponding to the question number being answered.

For example,
if you believe the correct answer to be option D, then your answer form should be marked as:

Q1 (A) (B) (C) ~~(D)~~ (E) (?)

(b) If you do not know the answer the answer to a question, pencil across the 'don't know cell "(?)"

For example,
if you do not know the answer to question 3 then your answer form should be marked as:

Q3 (A) (B) (C) (D) (E) ~~(?)~~

In order that this assessment is a true reflection of your performance, you are requested not to guess the answer to any question where you do not know the answer.

(c) If you wish to change any of your answers, cross out any cells inadvertently pencilled across.

For example,
if you wish to change your selected option for Q1 from (D) to (B) your answer form should be marked as:

Q1 (A) ~~(B)~~ (C) ~~(D)~~ (E) (?)

1. The Ishikawa diagram can be used for:

Key A. Showing correlation
B. Showing cause and effect relationships
C. Showing process flow
D. Separating the Vital Few from the Trivial Many
E. For identifying driving and restraining forces

2. In Phase II of a 4 phase QFD, Design Characteristics are translated from:

Key A. Production controls
B. Customer wants
C. Substitute Quality Characteristics
D. Rates of importance
E. Process parameters

3. The performance of Halewood and Saarlouis plants are compared by considering the value of the process Capability index Cpk for a specific product characteristic. The value for Halewood is 1.6 and that for Saarlouis 2.0. Which process produces the higher quality?

Key A. Halewood
B. Saarlouis
C. Neither
D Both equal
E. Can't say without more information

4. In performing a designed experiment the following <u>should</u> always be used:

Key A. An orthogonal array
B. Control group
C. Noise factors
D. Regression analysis
E. A confirmation run

5. As part of the production planning phase of QFD which of the following techniques is appropriate?

Key A. Substitute Quality Characteristics
B. xbar/R Charts
C. Force field analysis
D. Customer surveys
E. Function tree

6. In communication which of the following inhibits effective listening?

Key A. Restating of a communication
B. Building on an idea
C. Evaluating the merit of an idea
D. Stating one's likes about a proposal
E. Stating one's concerns or wishes about a proposal

7. In the Kano Model of Quality Features, basic quality:

Key A. is spoken about by customers
B. is one-dimensional
C. is an excitement feature
D. can cause customer dissatisfaction if not fully achieved
E. will cause complete customer satisfaction if fully achieved

8. EAO, Mazda and NAAO have tolerances of $\pm$ 0.25mm, $\pm$ 0.70mm and $\pm$ 0.50mm respectively on an identical stamping characteristic. Who is likely to produce the highest quality in the longer term, measured in terms of this characteristic:

Key A. EAO
B. Mazda
C. NAAO
D. All equal
E. Can't say without more information

9. Design FMEA teams establish the detection rating on the basis of:
Key A. The likelihood of the FMEA team detecting a failure mode
B. The parts supplier Q1 status to detect a failure mode
C. Manufacturing plant ISO 9000 approval
D. The likelihood of the Design Verification Plan detecting a failure mode
E. The existence of a failure mode

10. A Pareto diagram can be used to:
Key A. Prioritise action
B. Control a process
C. Assess correlation
D. Show frequency distribution of variable data
E. Identify process flow

11. All eight disciplines of the TOPS(8D) problem solving process should be considered:
Key A. When the severity of the concern warrants a team effort
B. On all concerns discussed at the Quality Strategy Committee
C. Where multiple causes are present
D. On every concern with more than 5TGW/100
E. When the concern can be solved by an individual

12. Which of the following procedures is not part of the task element of a meeting?
Key A. Analysis
B. Team process
C. Problem solving
D. Planning
E. Decision making

13. FMEA is concerned primarily with:
Key A. Addressing negative quality
B. Measuring basic quality
C. Detecting positive quality
D. Upgrading performance quality
E. Providing excitement quality

14. Which of the following is an important element of Company Total Quality Excellence
Key A. A customer focus
B. Defect detection
C. Kanban
D. A DQR strategy
E. Volume production

15. In performing a 'Taguchi' style experiment which of the following is of prime concern?
Key A. Changing factor level
B. Minimising the use of energy
C. Using an orthogonal array
D. Minimising the effect of noise
E. Finding root cause

16. Maintenance during a team meeting is:
Key A. Keeping the team's timing and agenda on target
B. The business of the meeting
C. Observation of the meeting dynamics
D. Time spent paying attention to the feelings and relationships of team members
E. The expression of conflicting options during the meeting

17. For effective process ownership the process owner must:
Key A. Have knowledge of TOPS(8D)
B. Have authority to make changes
C. Be Management Roll
D. Have intimate knowledge of the process
E. Have responsibility for maintaining control charts

18. In which TOPS(8D) discipline would you expect to see the use of IS/IS NOT information?

Key A. D7 - Prevent recurrence
B. D5 - Choose and verify permanent corrective actions
C. D2 - Describe the problem
D. D6 - Implement permanent corrective actions
E. D3 - Implement and verify interim containment actions

19. During the Warm Down process at a meeting, the team should:

Key A. Prepare the agenda for the next meeting
B. Assign tasks to be completed before the next meeting
C. Prepare the minutes of the meeting
D. Establish an action plan
E. Acknowledge participants' contributions to the meeting

20. Which of the following techniques are appropriate to apply in an effort to reduce common cause variation in a manufacturing process?

Key A. Experimentation
B. TOPS(8D)
C. FMEA
D. SPC
E. QFD

21. Choose the statement which best demonstrates the advantages of using descriptive rather than evaluative feedback:

Key A. It lets people know when they have gone wrong or made a mistake
B. It gives the individual information about their behaviour without judgement and without evoking defensive feedback
C. It avoids having to commit oneself to a position of judgement about others, thereby avoiding conflict
D. Evaluative feedback doesn't give the individual the option of changing behaviour
E. Descriptive feedback avoids discussion and wasting time because responses from the other person are avoided

22. During the design deployment phase of QFD, which of the following is not appropriate?

Key A .Affinity diagram
B. Relationship matrix
C. Parameter design experiment
D. Process FMEA
E. Function tree

23. The selection of nominal values for robust project/processes is achieved through application of:

Key A. Taguchi's Loss Function
B. Taguchi's System Design
C. Taguchi's Parameter Design
D. Taguchi's Tolerance Design
E. Taguchi's energy transfer model

24. In a meeting, the facilitator should take main responsibility for:

Key A. The completion of the task
B. The management of team members' time
C. Team members' emotional and physical needs
D. Recording team members' views
E. Reporting team decisions

25. A process which is in control:

Key A. consistently produces parts to specification
B. is subject only to special cause variation
C. is subject only to common cause variation
D. is sampled every hour
E. has Cpk>1.33

Appendix F Perceptions of the Organisation Questionnaire

The following questions relate to your parts of the Company . For example, if you work in Company of Britain, please relate your answers to your experiences with Company of Britain.

Please answer the question by circling the appropriate number. For example :

The Company is a great place to work.

Not at all			Somewhat					Completely
1	2	3	4	5	6	7	8	9

This would mean that you think the Company is a great place to work.

All answers are confidential. Please answer honestly. There are three sections :

Perceptions of the Organisation

1. Management at Company is sincere in its attempts to meet the employees point of view.

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

2. There is a strong sense of community, a feeling of shared interest and purpose among the managers of Company .

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

3. Our management would be quite prepared to gain advantage by deceiving the employees.

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

4. I feel quite confident that the Company will always treat me fairly

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

5. There is little conflict between managers

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

6. Meaningful co-operation and innovations in the Company are stifled because of too many vested interests

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

7. Managers are more interested in achieving the organisational goals than in personal advancement

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

8. Company has a poor future unless it can attract better managers

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

9. Sometimes I worry that asking for help at work might look like I can't do my job.

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

10. Asking for help from my colleagues can sometimes be humiliating

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

11. I am wary of asking for help publicly at work

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

12. It takes courage to ask for help in this organisation

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

13. I have to be careful when I talk to colleagues about work difficulties

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

14. Asking someone for help at work is as easy as asking a favour from a friend

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

15. I am happy to admit it when I need help to do my work

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

16. It is expected that one asks colleagues for help at work

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

17. Management at work seems to do an efficient job

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

18. Management can be trusted to make sensible decisions for the future of Company

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

19. There is considerably more competition than co-operation among the managers in Company .

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

20. I have full confidence in the skills of my colleagues

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

21. Most of my colleagues can be relied upon to do as they say they will do

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

22. I can rely on my colleagues not to make my job more difficult by careless work

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

Appendix G Perceptions of Quality (Ipsative) Questionnaire

Below are a set of pairs of statements, A and B. Your task is to choose the one which you agree with most.

Although in some cases you may feel you want to choose both or neither, or somewhere in between, please make a choice by circling A or B for every question.

1. A. To improve product quality, more effort should be directed towards preventing problems

   B. To improve product quality, more effort should be directed towards creating innovative products

2. A. To improve product quality more effort should be directed towards solving problems

   B. To improve product quality more effort should be directed towards creating innovative products

3. A. Quality is about exciting the customer

   B. Quality is about preventing mistakes

4. A. Problem solving should be recognised and rewarded in the same way as problem solving

   B. Innovations should be recognised and rewarded in the same way as problem solving

5. A. Quality products must have new and exciting features

   B. Customers prefer new products to be reassuringly familiar

6. A. The quality of a product or service depends entirely on our technical ability

   B. To make significant improvements in quality you need to be creative in your solutions

7. A. An effective measure of quality is a customer satisfaction index

   B. An effective measure of quality is the number of customer complaints

Appendix H Perceptions of Quality (Normative) Questionnaire

Below are a series of statements. Please indicate how much you disagree or agree with the statement by circling a number.

For example, circling '1' would mean you strongly disagree and circling '9' would mean you strongly agree.

1. Customers don't know what they want until experts show them

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

2. Sometimes we should give people what we think they need, and not what they say they want

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

3. The customer is very often wrong

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

4. When evaluating quality it is always better to consider the opinions of internal specialists rather than external clients

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

5. My contribution to the overall product is negligible

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

6. Most new features on our products are merely gimmicks

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

7. It is important to take time out to develop new and exciting concepts

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

8. To make significant improvements in quality you need to be creative in your solutions

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

9. I ought to have a sound understanding of the skills in other engineering areas

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

10. A quality product must have new and exciting features

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

11. Customers prefer new products to be reassuringly familiar

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

12. A good way to redesign a new product is to adapt the old design

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

13. Once you have a system, it is best to stick to it

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

14. Customer satisfaction is all that matters

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

15. When solving problems, I prefer to start with traditional approaches

Strongly Disagree	1	2	3	4	5	6	7	8	9	Strongly Agree

Appendix I Stakeholder Code Book

Degree of Involvement

Full time:
< 2 years/> 2 years
Attended modules:
Participation: self-initiated/other-initiated

Influence on Success

Passive /Champion
LOC internal/LOC external

Objectives of the programme

Quality improvement
Simultaneous engineering
Improve communication
Team building
Change work style
Attitude/behaviour change
Improve technical knowledge
Customer satisfaction/orientation
Overall same company work style

Need for the programme

Yes /No

Relevance to Daily Work

Tech. very/direct
Tech. relev/indirect
Behav. very/direct
Behav. relev/indirect

Relevance of Behavioural Skills to Engineers

Very important/Important/Little importance

Balance of Skills

Right/Wrong
Tech. > Behav./Tech. = Behav./Tech. < Behav.

Sequence

Yes /No/Cherry-picking

Preference for Group Mix

Cross-functional/Departmental/Ages mixed

Transfer of Training

Technical skills
Behavioural skills
No transfer
Use already/no difference
Need critical mass

Influence on Success of Company

Quality improvement
Simultaneous engineering
Economic effects
Competitiveness
Culture change -
Customer satisfaction/orientation

Barriers to Success

Time for training/pressure for release
Time for transfer
Resistance to change
Ideological
Critical mass
Transfer/consultation
Internal economic reasons
Timing for implementation
Lack of management support

Programme relationship to TQE

Partial link/methodology
Identical
Unknown/unsure

Change Attitude to Quality

Yes/No

Change in Practice

Yes/No

Customer/Supplier Impact

Train suppliers
Yes /No
Train other functions
Yes/No
Improved communication
Yes/No

Organisational Climate

Reasons:
Internal economic/External economic
Not so bad/improving
De-motivating/frustrating
Unsatisfactory/low morale
Bad
Insecure
Fear/nervous
Much pressure

Organisational Culture

Traditional style management
Autocratic
Bureaucratic
Team focused development

Programme Change Climate

Yes/No

Climate Influencing Success the programme

Yes/No

Management Support

Top yes/Top no
Middle/low yes/Middle/low no

Type of Support

Passive/Practice what they preach/Financial

Perception of Education and Training

Negative/Positive
Programme altered perception
Yes/No

Future of the programme

Continuous/Discontinued
Repeat/Not repeat
Flavour of the month/fad
Yes/No

Future of Company

Successful/more competitive
/Less competitive/Unsure

Programme influence on Future of Company

Yes/No

Programme as Training Intervention

Different/Not different
Internal trainers
Greater credibility
Content
Org. framework

European Context

Improved communication
Differences in training material
Content
Practical. influence