University of Nottingham

Division of Manufacturing Engineering and Operations Management

Structured Evaluation of Training in Virtual Environments

by

Mirabelle D'Cruz, BSc (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, June 1999

Table of Contents

- *Report A: Design of Questionnaires and Observation Forms*
- *Report B: Consent Form, Pre-questionnaire, Observation Form during training part 1/2. Observation Form during training part 2/2, Observation Form during performance and Post-questionnaire*
- *Report C: Breakdown of Results*

Appendix (IV) **Further Information for chapter eight**

- *Report A: Design of Questionnaires*
- *Report B: Consent Form, Pre-questionnaire, Feedback Forms, Post-questionnaire and Comparison of methods questionnaire*
- *Report C: Breakdown of Results*

Acknowledgements

I would like to express my thanks to Professor John Wilson for all of his guidance during the course of my Ph.D.

I also extend my thanks to the Virtual Reality Applications Research Team (VIRART) and other members of the Manufacturing Engineering and Operations Management department for their friendship and continuous support throughout the past few years - especially Richard Eastgate, Maria Gilgeous, Sue Cobb, Sarah Nichols, Amanda Ramsey, Joanna Crosier, Steven Kerr, Victor Bayon, Helen Neale and Lynne Mills.

Furthermore to my family - Mum, Dad, Michelle, Reuben, Justin, Scott, Pam, Chris and Becca - who were always there as a constant source of support and encouragement.

Finally my deepest thanks and all of my love is reserved for Rupert, my husband, who always looks after me and knows the right things to say.

List of Figures and Tables

Figures

Tables

Glossary of Terms

Many of the following terms are used throughout this thesis. The definitions are explained in greater detail in the chapters but are noted here also for quick reference.

- **Virtual Reality (VR) -** The combination of systems that are used to create and maintain virtual environments
- **Virtual Environments (VEs) -** computer-generated representations of real or imaginary environments, experienced as three dimensional via a number of sensory channels. Objects within these environments are independent of the user and can display real world behaviour. The user has autonomous control - the freedom to navigate and interact with the objects. This interaction occurs in real-time and the users experience feelings of presence and/or involvement.
- **Input device -** device such as a mouse or joystick, which allows the user to interact with the VE.
- **Desktop VR -** the most basic of the VR systems which consists mainly of a standard affordable computer with specialised software.
- **Projected VR** - VR systems which use large projection displays on one or more surfaces
- **Artificial Reality or Video Mapping** - VR systems which merge a video image of the user with computer graphics.
- **Augmented Reality -** VR systems which use transparent displays to allow simultaneous viewing of the real world and information produced by a computer.
- **Headset Systems -** VR systems which consist of a head mounted device (HMD) which the user wears to view and hear the VE.
- **VET applications -** acronym for Virtual Environment Training Applications

Abstract

Virtual Environments (VEs) created through Virtual Reality (VR) technologies have been suggested as potentially beneficial for a number of applications. However a review of VEs and VR has highlighted the main barriers to implementation as: current technological limitations; usability issues with various systems; a lack of real applications; and therefore little proven value of use. These barriers suggest that industry would benefit from some structured guidance for developing effective VEs. To examine this 'training' was chosen to be explored, as it has been suggested as a potential early use of VEs and is of importance to many sectors.

A review of existing case studies on VE training applications (VETs) examined type of training applications and VR systems being considered; state of development of these applications and results of any evaluation studies. In light of these case studies, it was possible to focus this work on the structured evaluation of training psycho-motor skills using VEs created by desktop VR.

In order to perform structured evaluation, existing theories of training and evaluation were also reviewed. Using these theories, a framework for developing VETs was suggested. Applying this framework, two VETs were proposed, specified, developed and evaluated. Conclusions of this work highlighted the many areas in the development process of an effective VET that still need addressing. In particular, in the proposal stage, it is necessary to provide some guidance on the appropriateness of VET for particular tasks. In the specification and building stages, standard formats and techniques are required in order to guide the VE developer(s) in producing an effective VET. Finally in the evaluation stage, there are still tools required that highlight the benefits of VET and many more evaluation studies needed to contribute information back to the development process. Therefore VEs are still in their early stages and this work unifies existing work in the area specifically on training and highlights the gaps that need to be addressed before widespread implementation.

Chapter 1: Introduction

1.1 Background to this Research

‘Virtual Reality (VR)’ is a computer technology that creates ‘Virtual Environments (VEs)', in which the users have the freedom or ‘autonomy’ to ‘interact’ with ‘virtual objects’ using a number of their senses through various 'input' devices. This interaction occurs in ‘real-time’ and the users experience feelings of ‘immersion’ and/or ‘presence’. (The features of this definition are discussed in Chapter Two). During 1993 - 1994, the Virtual Reality Applications Research Team (VIRART) in the Department of Manufacturing Engineering and Operations Management at the University of Nottingham, conducted a feasibility study funded by the Engineering and Physical Sciences Research Council - EPSRC (Grant GR/J57643). This one year study was entitled “Applications of Virtual Reality in UK Manufacturing Industry: A Feasibility Study” of which the author was one of the principle investigators on the grant.

The main aims of the study were to provide information on: industry’s understanding and perception of Virtual Reality (VR); their potential applications; their readiness for implementation; and the identification of their needs. The investigation involved much in-depth work with users and potential user companies e.g. Rover, British Telecom, Crossrail, Vauxhall, Ford, Rolls Royce and Associates, and close contact with suppliers and developers of Virtual Reality systems e.g. Superscape, Division, Virtuality and Virtual Presence. (Subsequent reports from this work are listed under the references as Cobb et al 1994a, 1994b, 1994c, 1994d; and D’Cruz et al, 1994, Wilson et al, 1994). This work was part of an overall initiative called the MOVE (Manufacturing Operations in Virtual Environments) programme, which was part of VIRART’s commitment to identifying and addressing the needs of manufacturing industries (Wilson et al, 1995). A revision of this information was disseminated in Wilson et al, 1995; Wilson et al, 1996 and is summarised in Chapter Three of this thesis.)

In particular, the findings of the study identified a number of major barriers to implementation. These were technological limitations, usability issues, the lack of examples of applications and the lack of any real evidence of benefits (Wilson et al, 1994; Wilson et al, 1996 p.118). These barriers were also confirmed by other reports by independent consultants (Leston, 1994; Frost and Sullivan, 1996) as well as, national bodies such as, the U.S. National Research Council (NRC, 1995). The main barrier still is the technological limitations imposed by current systems which in turn create many of the usability problems. There is a substantial gap that exists between the technology available and the technology required to fulfil the potential of the applications being envisaged for Virtual Environments (VEs). This technology gap however is constantly changing as much work is concentrated on developing better, more usable systems. However the lack of examples of 'working' applications of VEs and therefore the lack of any real evidence of benefits, are the issues that will be studied in this thesis.

1.2 Definition of the Problem

Virtual Environments (VEs) are still considered a relatively new concept with potential benefits for industrial applications. Among these applications, it has often been suggested that training may be a particularly "*powerful and useful early application*" (NRC, 1995). However there is still little evidence to support this, partly due to the lack of existing applications and the lack of evaluation studies to support evidence of real benefits and transfer. Examples of VE training (VET) applications tend to be demonstrations of 'proof-of-concept' rather than 'working' applications that can be measured for effectiveness. Generally the main examples of VETs have been funded by the military and aerospace sectors (Johnston, RS. 1987; Moshell et al, 1990; Alluisi, EA. 1991; Levison and Pew, 1993; Magee, 1993; Bowen Loftin and Kenney, 1994; Kenney and Saito, 1994; Bowen Loftin, 1995; Cater & Huffman, 1995; Johnson and Wightman, 1995; Zeltzer, 1995; Zeltzer et al, 1995; Zeltzer and Pioch 1996; McLin and Chung, 1996). However these applications have used 'high-end' VR systems which are beyond the scope of most other industrial sectors. Also their studies have shown that these VR systems have resulted in usability problems and side effects which have effected performance of the tasks in the VE and subsequently in the real world too. Therefore it is difficult to relate their experiences with VEs to the immediate requirements of industrial training, as it has been indicated that early use of VEs in industry is likely to be with low end ‘desktop' VR systems or ‘off-head displays’ (NCR, 1995; Wilson et al, 1996). Therefore the concentration of this research is specifically on training applications developed through ‘desktop' VR systems.

Of the VET applications that have been reported by industry (CyberEdge Journal, 1994; Wittenberg, 1995; Adams, 1996; Morrissey, 1996; VR News, 1996a; VR News, 1996b; Wilson et al, 1996; I/S Analyzer, 1997; VRET, 1997) generally evaluation of these applications are only really now being considered. However, they are likely to find the process highly problematic when considering the difficulties found by researchers in the area (Regian et al, 1992; Kozak et al, 1993; Kenyon and Afenya, 1995; Witmer et al, 1996). The problems lie with the large number of factors to be considered, especially as a VR system can have a number of different configurations, a VE can have a number of different designs and the whole training process is dependent on the individual characteristics of the trainee, the training content and the most appropriate way of learning that content. These are all areas which need careful consideration before being able to develop effective VET applications. Therefore while the focus of this research is on evaluating VET, part of the work will also consider the influence of the development process.

1.3 Research Aims and Objectives

Considering the current situation of VET in industry, the main aim of this research is as follows:

To investigate and further the development and evaluation process of Virtual Environments (VEs) in the field of industrial training.

This aim will be achieved through the following objectives:

1) Critically review the capabilities and limitations of current VR systems and VEs for their impact on training applications;

2) Identify the types of training applications currently being pursued by industry using VEs, highlight the findings and major problems that need addressing;

3) Investigate existing theories and methodologies of training and evaluation for their recommendations on developing and evaluating effective VET applications;

4) Suggest a structured framework in which to develop and evaluate effective VET applications;

5) Develop a VET application based on the needs of industry and explore the costs and benefits against other forms of training, through experimental work;

6) Provide recommendations for the future of VET.

1.4 Research Methodology

In order to investigate these objectives a number of varied methods were used in line with a scientific approach to research as outlined by Dane (1990). The methods are discussed as follows.

1.4.1 Literature Reviews and information gathering

Literature searches continued throughout this study through the University of Nottingham and the British Library via the Libertas System, the Bath Information Database System (BIDS) and the Psychology Literature (PsycLit) CD-ROM. Further up-to-date information was gained via the internet resources through various web-sites and newsgroups including: sci-virtual world's news group, ukvrsig mailing list and vrml mailing list.

Information was gathered through contacts with the main developers and suppliers of VR systems such as: Superscape User Group Meetings; Division; Virtuality; Virtual Presence; and the main reporters on VR - Mike Bevan, editor of VR News and Ben Delaney, editor of the former CyberEdge Journal.

Presentation at and attendance of some of the main VR and training conferences also aided information gathering and contact with companies for case-studies. These conferences included: Frameworks In Virtual Environments (FIVE) 1995, London, UK; Virtual Reality World 1996, Stuttgart, Germany; The World Open Learning for Business Conference 1996, Wembley Conference Centre, London, UK; Virtual Reality Universe (VRU) 1997, San Jose, US; and Virtual Reality Education and Training (VRET) 1997, Loughborough, UK.

1.4.2 Archival Research

During the data collection for the MOVE programme (discussed above in section 1.1) information was gathered from a number of industrial sectors as well as, manufacturing. While the focus of MOVE was on manufacturing, for this study the data was re-analysed to include all the other industrial sectors in order to provide a more general overview of industrial perceptions.

1.4.3 Field Research

A number of leading research groups and companies in both the areas of VEs and training were contacted or visited and informal interviews were carried out. These groups included: Westland Systems Assessment Laboratory (WSAL), Somerset; NCR, Dundee; Rolls Royce and Associates, Derby; the Human Interface Technology Laboratory (HITLab), Seattle; Adams Consulting, Chicago; and Motorola, Chicago. As well as, multimedia and computer-based training (CBT) companies such as: CBL Technology Limited, Derby; Logicom; and TV-i, London.

1.4.4 Experimental Research

Given the outcomes of the information gathered from the literature, archive data and fieldwork, methods within a framework were produced to guide the development and evaluation of VET. Two experiments were designed to explore aspects of VET in detail and to measure it against other training methods.

1.5 Contribution of Research to the Academic Community

This research unifies existing knowledge on the capabilities and limitations of VR systems and VEs, with particular consideration of the needs of industrial training. It reviews existing theories and methodologies of the training development process in view of its contribution to the development process of effective VET applications. It then provides an insight into the potential value of VET through experimental work.

In summary this research is intended to further previous studies in virtual environment training by concentrating on the immediate needs of industry - desktop VR and training of basic procedural skills.

1.6 Contribution of Research to Industry

This research is focussed on training which is of importance across many different industries. Therefore this research has relevance to many companies. Also the direction of the research has been dictated by the needs of industry. It began through an examination of information provided by industry through a number of surveys and in-depth interviews of their needs and requirements of VEs. Through these, desktop VR and training have been identified as suitable for matching their needs and VET using desktop VR has been examined in this study. In summary this research attempts to address some of the immediate VE needs of industry that they have expressed as their requirements.

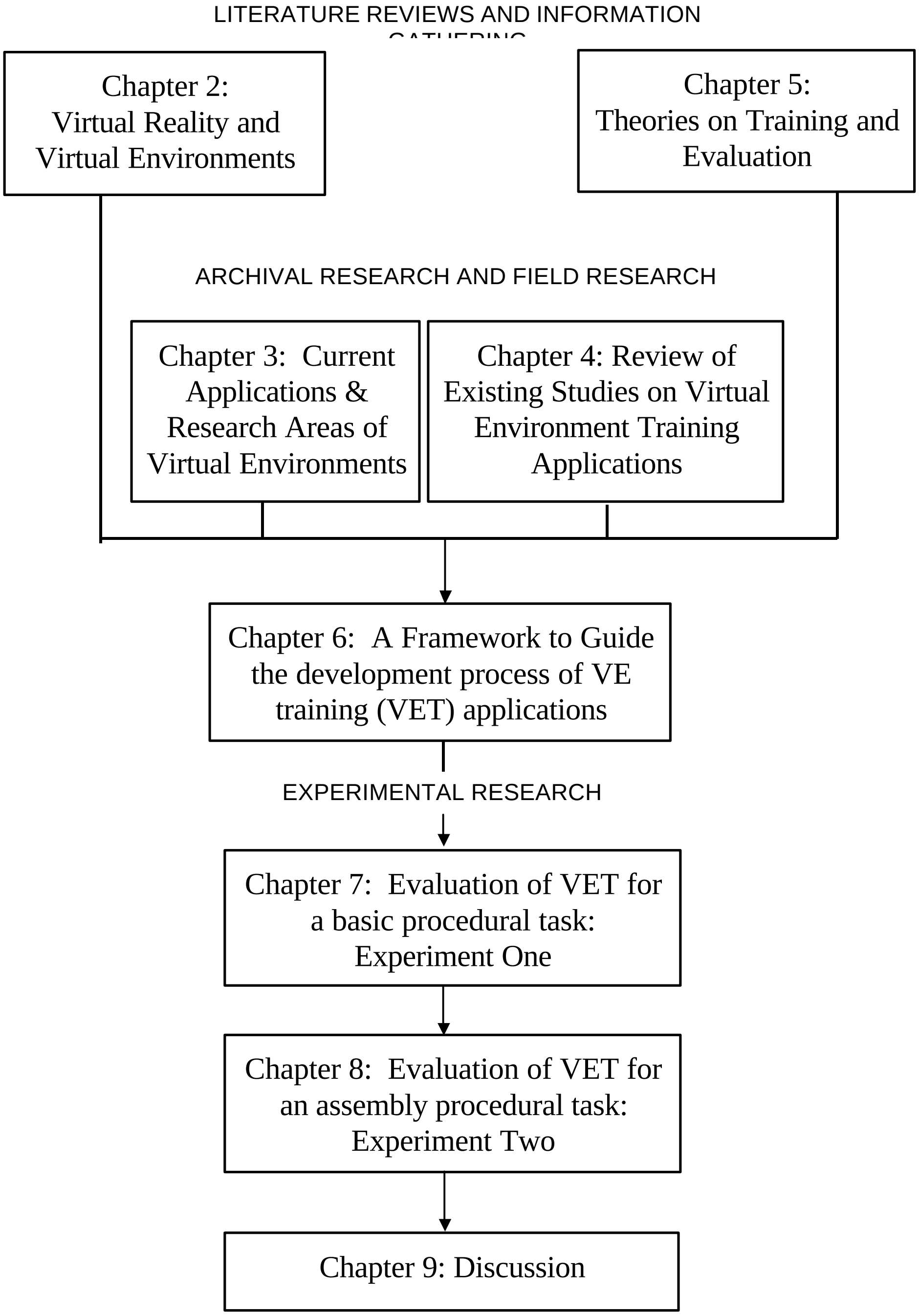
1.7 Structure of thesis

The structure of the thesis is illustrated in Figure 1.1. Chapters two and five are presentations of information gathered through literature, archives and field research. Chapters three and four are a combination of reviews of other work, as well as, surveys, interviews, visits to companies and conferences which have been part of the author's own investigations. Chapters six, seven and eight consist of the author's own theories and experimental work supported by the previous chapters and chapter nine is a discussion of the findings of the thesis and recommendations for future research.

In terms of content, chapter two provides some general information about Virtual Reality (VR) and Virtual Environments (VEs). This information is to provide an understanding of the existing area and includes definitions, historical background and a technical overview. Following on from this, chapter three presents details of VIRART's MOVE programme which provided some understanding of industry's perceptions, needs and requirements of VR and VEs and then a brief look at the potential applications that are currently being explored. Then chapter four reviews a number of existing case-studies specifically on training from the military, aerospace and industry, as well as, a look at some of the experimental work that has already been carried out. The discussion of this review identifies the types of training tasks - navigation and procedural - and some of the positive and negative aspects of the studies. Chapter five provides a brief overview of existing information from the training field which has direct relevance to this work. Lessons from previous similar technologies such as teaching machines, simulators, computer based training and multimedia provide an understanding of the difficulties faced when evaluating VEs.

These previous chapters provide the information required to develop a framework described in chapter six which includes the development process of VET. In order to provide a structured approach to evaluating VET applications as they are very much inter-linked. Such a process has currently not been identified. Therefore given the experience with the MOVE programme, visits to a number of companies and working with companies in developing their VET applications, this framework suggests a process for developing effective VET applications. This framework identifies the areas of proposing, specifying, building and evaluating VET applications. It shows that while the first three of these areas can be mostly supported by existing theories and literature, the area of evaluation has a number of issues specific to VEs which need to be examined further. In order to do this chapters seven and eight describe experimental work carried out to examine these issues more closely. These experiments apply the evaluation methods discussed in chapter six on different VET applications. The applications are also measured against examples of conventional training methods. Finally chapter nine concludes the thesis with an discussion of the main findings, the contribution of this work to the current status of VET and suggestions for the direction of further research.

Figure 1.1. Structure of the Thesis



CHAPTER 2: Virtual Reality (VR) and Virtual Environments (VEs)

2.1 Introduction

This chapter covers the overall area of Virtual Reality (VR) and Virtual Environments (VEs). It begins with definitions in section 2.2 and leads on to a brief history of VR in section 2.3. This is followed by a summary of the current state-of-the-art of the technology and a discussion about some of the main points that contribute to this research.

2.2 Definitions

2.2.1 Virtual Reality (VR) and Virtual Environments (VEs)

At present there are no single concise or generally accepted definitions of Virtual Reality (VR) and Virtual Environments (VEs). This is partly due to the continual state of evolution of the many technologies involved and also due to people using the terms to mean a variety of things. Some of these definitions are discussed next. However in the context of this work the term Virtual Reality (VR) is used to refer to:

The combination of systems that are used to create and maintain virtual environments.

In the context of VR, the attributes of virtual environments (VEs) are listed as follows (based on Wilson et al, 1996; p. 4):

- *computer-generated representations of real or imaginary environments;*
- *experienced as three dimensional via a number of senses - visual, aural and/or tactile;*
- *objects within these environments are independent of the user and can display real world behaviour;*
- *the user or users have autonomous control - the freedom to navigate and interact with objects, using a number of different viewpoints;*
- *interaction occurs in real-time; and*
- *the users experience feelings of presence and/or immersion.*

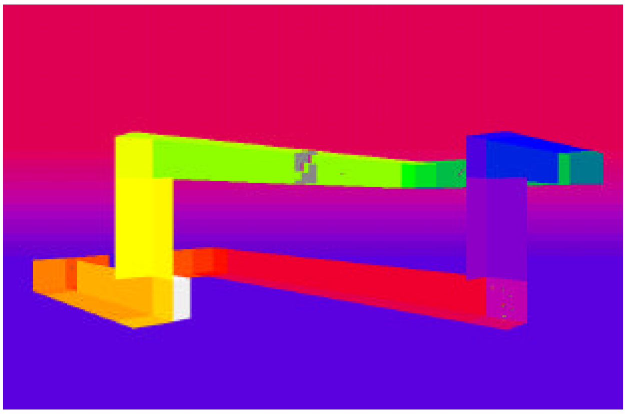
VR is a new level of human computer interaction where, in principle, people are able to perform the activities which they are best suited e.g. logical reasoning, problem-solving, interpreting information in different ways to form different conclusions and so on. Computers are able to perform the activities that they are suited to - storing a huge amount of information, performing calculations at great speed and accuracy and displaying the information in different ways.

VR systems create VEs which are computer-generated simulations of real or imaginary worlds. They can represent an existing environment or one that does not yet exist or environments that are not possible for us to see or even ones that can only exist in the imagination (see figures 2.1, 2.2 and 2.3 overleaf).

Figure 2.1: A representation of an existing environment. This is the showroom of NCR in Glasgow

Figure 2.2: A representation of an environment which does not yet exist. This shows a virtual environment of the new VIRART office before the actual furniture had arrived.

Figure 2.3: An imaginary environment. This maze was developed by VIRART to test people's ability to use different input devices.



These environments are experienced as three-dimensional through a number of senses - sight, sound, touch etc., through a number of technologies called ‘input’ devices (which are discussed in section 2.4). The objects within these environments are independent of the user. They can display real world behaviour or autonomy in terms of gravity, acceleration or friction so that they behave appropriately when interaction occurs with another object or user. The user or users within the environment have a certain amount of autonomous control which provides the freedom to navigate and interact with the objects, using a number of different viewpoints. These viewpoints can either be as a person or an object or a ‘ghost’ (the ability to move in an environment without any restrictions at all). This can all occur in ‘real-time’, that is, the computer can generate the image of the VE many times a second so that the correct view (according to the behaviour of the user) is maintained in continuous motion, as it would appear in the real world.

Finally the user is said to experience feelings of immersion and/or ‘presence’ in the VE. Slater and Wilbur (1995) distinguish between immersion and presence by defining ‘immersion’ as “*the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant,” (p. 13)* and presence as “*a state of consciousness, the (psychological) sense of being in the VE” (p. 14).* These two attributes have been the focus of much research (Heeter, 1992; Held and Durlach 1992; Loomis, 1992; Sheridan, 1992; Zeltzer, 1992; Barfield and Weghorst, 1993; Slater et al, 1994; Witmer and Singer, 1994 and Barfield et al, 1995) as it is felt that, along with ‘interactivity’ they distinguish VEs over other similar technologies.

2.2.2 Other definitions

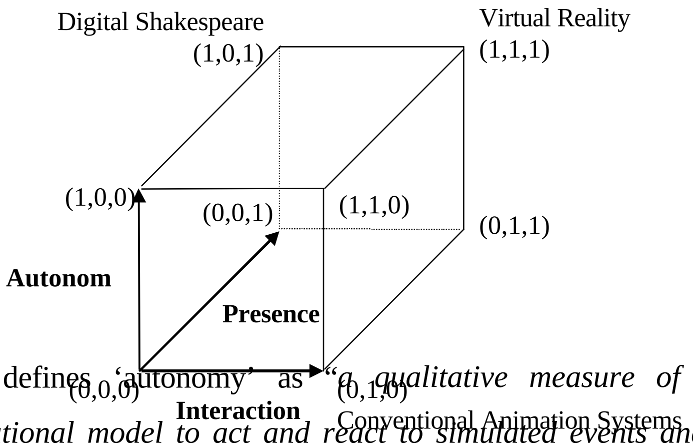
The difficulty in defining VR and VEs has occurred because of a number of problems. Firstly, many different disciplines from art, psychology, design, computer science, engineering, manufacturing, medicine etc., have become very interested in applications using VEs. However these diverse disciplines can cause confusion with terminology as in some cases the same words may have different definitions making it difficult for these disciplines to communicate effectively. The National Research Council (NRC, 1995) provided a good example of this, "*computer scientists naturally use the terms* input *and* output *in reference to the computer, psychologists use these terms in reference to the human user. Thus, in a virtual environment system, what is output to the psychologist is input to the computer scientist."* (P. 14). However because VR is a computer technology which focuses on allowing a persons ability to interact with information presented by the computer, the future development of VEs relies on collaborative work between the many disciplines.

There is also the problem of misleading reports about the technology. In the attempt to 'jump on the bandwagon' a large number of conferences, magazines and publications began in the early 1990's; companies were advertising their 'active use' of VR systems and even a few films were made - "Lawnmower Man", "Lawnmower Man 2" and "Disclosure". However, this resulted in false expectations of the technology which caused much disappointment to the point where researchers working with VR systems began using alternative terms like 'VE systems' and 'interactive environments', in order to avoid the confusion.

Also there is the problem in distinguishing them from other similar technologies. VEs have grown out of many technologies and is part of what people have termed 'graphic simulation systems' (Zeltzer, 1992), 'technologically mediated experiences or synthetic experiences' (Robinett, 1992), or 'synthetic environment systems' (National Research Council, 1995).

Zeltzer (1992) was one of the first authors to attempt a classification of VEs. His definition assumed that any VE has three components. Firstly, a set of models/objects or processes; secondly, a means of modifying the states of these models; and finally, a range of sensory modalities to allow the user to experience the VE. Zeltzer represents these components on a cube with scales relating to ‘autonomy’, ‘interaction’ and ‘presence’, as shown in Figure 2.4.

Figure 2.4: Zeltzer’s Autonomy, Interactions and Presence (AIP) cube (Zeltzer 1992, p.129)



Zeltzer defines ‘autonomy’ as “*a qualitative measure of the ability of a computational model to act and react to simulated events and stimuli, ranging from ‘0’ for the passive geometric model to ‘1’ for the most sophisticated, physically based virtual agent” (p. 127).* ‘Interaction’ is, “*the degree of access to model parameters at runtime (i.e., the ability to define and modify states of a model with immediate response). The range is from 0 for ‘batch’ processing in which no interaction at runtime is possible, to 1 for comprehensive, real-time access to all model parameters” (p. 127).* And ‘presence’ is “*our sense of being in and of the world ... engendered by our ability to affect the world through touch, gesture, voice, etc*.” Therefore the presence axis provides a crude “*measure of the number and fidelity of available sensory input and output channels” (p. 128)* which implies that the more the user’s senses are engaged by the environment, the more ‘presence’ they will feel. However the degree of sensory input and output is very much dependent on the application and the design of the VE. This design is guided by ‘selective fidelity’. (Johnston 1987, cited by Zeltzer 1992 and Robinett 1992). Generally it is impossible to reproduce the real world in huge detail and complexity, therefore depending on the application, the ‘sensory’ cues which are necessary for the user to fulfil the application must be carefully identified. Then the design of the environment must match as closely as possible the human perceptual and motor performance required for successfully completing the application. Given the limitations on current technology many trade-offs have to be made and how these effect the degree of presence has been the focus of much research work over the years.

Zeltzer's cube represents at the point (0,0,0) the early graphic systems e.g. graph plotters and charts, which have no autonomy, interactivity or presence, to the ideal VR system (1,1,1) with full autonomy, interaction and presence. However, Zeltzer suggests that current VR systems are at the (0,1,1) point, that is, a high degree of interactivity and possibly, presence but very limited degree of automation of objects.

Robinett (1992) proposed a taxonomy on 'synthetic experience' which he defined as "*perceiving a representation or simulacrum of something physically real rather than the thing itself*." (p.230). It considers experiences which rely on some form of technology e.g. computers and digital electronics, to interpret the user's actions and provide a response. Therefore it includes technologies like the telescope, microscope, television and telephone, as well as, teleoperation, VR and flight simulation. The systems are classified into nine categories each with further sub-divisions. The aim is to provide some clarity on the similarities and differences of the various 'technologically mediated experiences'. The nine dimensions and their possibilities are shown in Table 2.1 (overleaf). In Robinett's discussion of this taxonomy, he compares it with Zeltzer's model and Naimark's taxonomy on methods for recording and reproducing experience (Naimark, 1991). In general he states that his taxonomy attempts to cover the overall domains of both models for recorded, simulated and transmitted experience. He suggests that it offers a 'starting-point' for discussion, and, in particular, explores the potential possibilities for VR systems and their relationship to other systems.

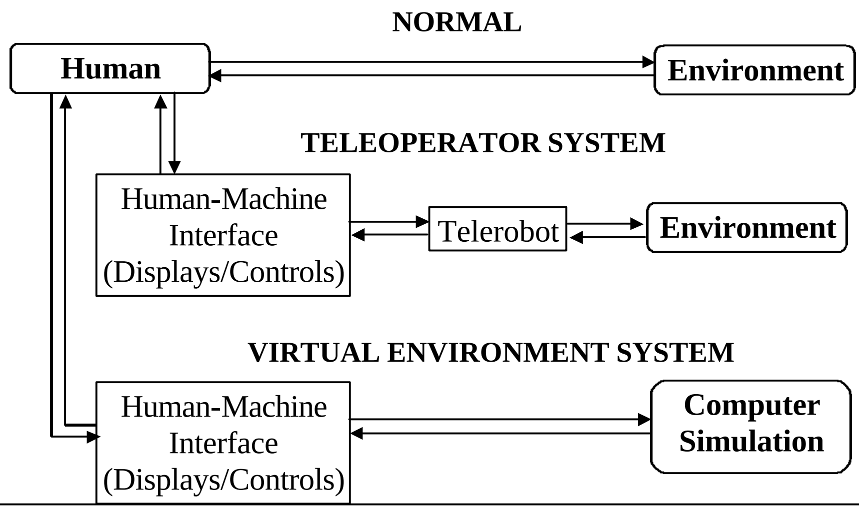
Table 2.1: Robinett's classification system for types of synthetic experience (Robinett 1992, p. 233)

Dimension	Possibilities	Examples
Causality	• Simulated • Recorded • Transmitted	• Flight simulator • Film • Teleoperation
Model Source	• Scanned • Constructed • Computed • Edited	• Night vision goggles • Video game • Computational Fluid Dynamics • Film
Time	• 1-to-1 • Accelerated (or retarded) • Frozen • Distorted	• Film • Time-lapse Photography • Photograph • Edited video recording of event
Space	• Registered • Remote • Miniaturised (or enlarged) • Distorted	• Night vision goggles • Teleoperation • Microteleoperation (Scanning-tunnelling Microscope) • Scanning-tunnelling Microscope with heights exaggerated
Superposition	• Merged • Isolated	• Augmented Reality • Virtual Reality
Display Type	• HMD • Screen • Speaker	• Virtual Reality • Video game • Recorded music
Sensor Type	• Photomultiplier • Scanning-tunnelling Microscope • Ultrasound	• Night vision goggles • Microteleoperation • Medical "X-ray vision"

Action Measurement Type	• Tracker and glove • Joystick • Force Feedback Arm	• Virtual Reality • Video game • Teleoperation
Actuator Type	• Robot arm • Scanning-tunnelling Microscope tip • Aircraft flaps	• Teleoperation • Microteleoperation • Remote piloted aircraft

More recently, the National Research Council (U.S. committee consisting of the National Academy of Sciences, the National Academy of Engineering and the Institute of Medicine) produced a report on ‘Virtual Reality - Research and Developments’ (NRC, 1995). They have suggested that VR systems are part of ‘synthetic environment (SE) systems’. That is, a system where “*the human operator is transported into a new interactive environment by means of devices that display signals to the operator’s sense organs and devices that sense various actions of the operator*” (NRC, 1995; p. 13). Other systems which belong to this group include ‘teleoperation’ and ‘simulator systems’. The difference between teleoperation and VEs was illustrated with the following diagram shown in Figure 2.5.

Figure 2.5: Schematic outline comparing a teleoperator system, a VE system and an unmediated (normal) system. (NRC, 1995; p. 17)



The figure shows that in a normal situation the ‘human operator’ and the environment, directly interact with each other. In a teleoperator system, the human operator interacts with the environment via some form of ‘human-machine interface’ and a ‘telerobot’. Sheridan (1992) defines a teleoperator as “*a machine that extends a person’s sensing and/or manipulating capability to a location remote from that person” (p.4).* He defines a telerobot as “*an advanced form of teleoperator the behaviour of which a human operator supervises through a computer intermediary. That is, the operator intermittently communicates to a computer information about goals, constraints, plans, contingencies, assumptions, suggestions and orders relative to a remote task, getting back integrated information about accomplishments, difficulties and concerns and (as requested) raw sensory data. The subordinate telerobot executes the task on the basis of information received from the human operator plus its own artificial sensing and intelligence” (p.4).* Therefore these systems are used in areas that may be too remote or dangerous for a human operator, in order to change the state of the environment. These systems have already been used in various applications and being developed for others including space, undersea oil and science, nuclear power plants, toxic waste clean up, construction, agriculture, mining, warehousing and mail delivery, firefighting and lifesaving, policing, military operations, assisted devices for the disabled, telediagnosis, telesurgery and entertainment (Sheridan, 1992; Chapter 2). However, in a VE system, the human operator interacts with the computer environment, so it could be said that the main purpose is to change the state of the human operator or the information stored in the computer. Therefore, the main differences between teleoperators and VEs are the purpose of the systems and the equipment used.

A more difficult distinction to be made is that of VEs and simulation systems. However the National Research Council (NRC, 1995) provide a clear list of where the term VE should be used rather than simulator, as follows:

- *the system is easily re-configurable by changes in the software*

- *the system can be used to create highly unnatural environments as well as a wide variety of natural ones*
- *the system is highly interactive and adaptive*
- *the system makes use of a wide variety of human sensing modalities and human sensorimotor systems and*
- *the user become highly immersed in the computer-synthesized environment and experiences a strong sense of presence in the artificial environment (p.22).*

Furthermore, they suggest that the focus of simulators and VEs are very different. In a simulator, it is usually the equipment e.g. aeroplane, vehicle, etc., which is the focus of the application and so physical mock-ups tend to be used and the 'simulated environment' represents distant locations. Currently, the focus of VEs is on the user and so the environment tends to represent the whole situation - near and far objects. Finally, the difference between other computer-based systems is considered mainly on the extent to which the system is 3D, interactive, multi-modal and immersive. These are recognised as the distinguishing features of VR systems.

2.3 A Brief History of Virtual Reality (VR)

There are numerous authors which have provided a history of VR (Fisher, 1990; Rheingold, 1991; Ellis, 1991a; Ellis, 1991b; Biocca, F. 1992; Gigante, 1993; Pimental and Teixeira, 1993; Burdea and Coiffet, 1994). A brief summary based on these authors is as follows: VR has developed through a combination of 'visions' and 'enabling' technologies. In the 1950s and 60s, people were developing the concept of a 'virtual reality'. The most famous of these 'visionaries' was Morton Heilig, a cinematographer in California who is recognised as the person who inspired much of the early industry with his 'Sensorama' machine (Heilig, 1992). This consisted of a motorcycle that a person sat on and watched video scenes of Brooklyn in the 1950s. It included sights, sounds, vibration of the cycle for feel, and even smells. However, it was not computer-based and not interactive.

It is generally accepted that VR, as recognised today, was begun by Ivan Sutherland, a researcher at Harvard University, with his paper “The Ultimate Display” (Sutherland, 1965). This reported the first ever head- mounted-device (HMD) or headset which could show computer-generated three dimensional images of cubes. At the same time, another researcher at the University of Wisconsin, Myron Krueger, was experimenting with what is now called, ‘artificial reality,’ (see section 2.4). His GLOWFLOW project involved a computer-controlled light and sound environment which responded to the behaviour of the users (Krueger, 1991). However their systems could not be made commercially with the technology which was then available. Also, VR was not considered a ‘scientific-enough’ subject, according to the American academic community and so information was difficult to disseminate.

During this time, there was a rapid development of ‘enabling’ technologies driven by various sources. These are technologies which ‘enable’ other technologies to exist. Small electromagnetic cathode ray tubes (CRTs) were produced by companies such as Thorn, Thomas and Hughes with military funding. A strong market for consumer electronics resulted in the production of small flat-panel displays and liquid crystal displays (LCD). These innovations allowed for smaller, lighter and cheaper display devices at a higher quality. Also, the personal computing revolution in the ‘70s and ‘80s made available fast, cheap, digital image generation, as well as high speed graphics workstations which could produce images faster and with resolutions of a higher degree. Tracking systems, which could translate the movement and orientation of a person in the real world into a computer environment were being developed (primarily by Honeywell and Polhemus again for military use). It was the convergence of all of these technologies which allowed affordable systems for applications.

In 1985, Michael McGreevy from NASA/Ames Aerospace, Human Factors Research Division, held an event to show his new ‘affordable’ head-mounted-device (HMD), made from currently available technologies. It used a motorcycle helmet with two LCD screens from a couple of mini televisions and a magnetic tracking device which connected the user to the computers which generated the images. It was the first suggestion that the technology was now affordable.

In the same year, Myron Krueger opened his ‘VIDEOPLACE’ in Connecticut Museum of Natural History. This used a combination of video camera, computer graphics and gesture/position-sensing technologies. The system allowed multiple users to interact with each other even though they were in different rooms.

In other places in the U.S., researchers were exploring glove technologies. Thomas Zimmerman is recognised as the inventor of the VR glove, (although in 1981, a researcher called Gary Grimes, who was working for Bell Laboratories, had patented a glove-based computer interface device. This glove used small switches at each finger joint to allow the user to interact with computer images. However Bell did not pursue the work (Rheingold, 1991)). Zimmerman’s glove was light-weight and used thin, pliable, hollow plastic tubes which conducted light. He developed it so that he could play true ‘air guitar’. When Zimmerman met Jaron Lanier (who was then famous for programming a video game called “Moondust” for Atari), they founded VPL Research Inc.

The real breakthrough for VR came when Scott Fisher of NASA/Ames commissioned VPL to develop a glove for their VR system. This reached the attention of the media with the result of a lot of ‘hype’ which raised the awareness of VR in the public domain. However, even though it is generally agreed that VR largely began in the U.S., the world’s first commercial VR system was launched in 1991 in the U.K. at Wembley, London, by Jonathan Waldern, a researcher at the former Leicester Polytechnic. He first designed the system in 1984 after reading a paper by Jim Clark, a student of Sutherland's in Utah. By 1988, he and some friends had built the first VR arcade system in his garage and after the launch it was placed in London's Covent Garden's Rock Garden Club and 'W' industries had begun the VR 'revolution'.

From 1991 until now - 1998, VR systems and VEs have been progressing their way along the learning and development curve. The technology is still placing limitations on what has been envisaged for VEs. However there is also still a need to understand how VEs can be used effectively. Research work in the area is now prolific around the world with Japan now becoming one of the major players along with the United States, the U.K. and Germany. A list of the research work being carried out around the world can be found in Appendix I. There are 20 countries actively exploring VEs, with numerous research sites which can be accessed via the internet. Information is now more readily available, companies are releasing their internal studies and researchers and industrialists are beginning to form collaborations over many different projects. The next ten years of VR and VEs looks set to be very interesting.

2.4 Current status of VR systems

There are a number of authors who have provided detailed overviews on VR technology such as Aukstakalnis and Blatner (1992), Earnshaw, Gigante and Jones (1993), Pimental and Teixeira (1993), Casey Larijani L. (1993) and Hollands (1997). However the continuous and rapid development of the technology implies that to get true up-to-date information the best sources are the various internet sites which are continually up-dated. The most useful sites which provide information, links to the suppliers and price lists, include:

(1) The 'HITLab's knowledge base project' co-ordinated by Toni Emerson, Director of Information Services at the Human Interface Technology Laboratory (HITLab), University of Washington, Seattle.

http://www.hitl.washington.edu/projects/knowledge-base/onthenet. html

(2) ‘Chris Hand’s VR stuff’, a Senior Lecturer in the Department of Computer Science at the University of De Montfort, Leicester and an active member of many VR groups.

http://www.cms.dmu.ac.uk/~cph/

(3) Ian Feldberg’s ‘VR buying list’, a senior member of staff at the Applied Physics Laboratory at the John-Hopkins University, Baltimore.

http://www.cs.jhu.edu/~feldberg/

(4) ‘Virtual Reality Resource List’ compiled by Shawn T. Rutledge, a software design engineer with Essential Wisdom, Arizona.

http://www.goodnet.com/~ecloud/

2.4.1 Types of systems available

The main types of VR systems tend to be based on the method that they interface to the user. The most common categories are ‘desktop VR’, ‘projected VR’, ‘artificial reality’, ‘augmented reality’ and ‘headset systems’. The components of each system - hardware, software, tracking systems, visual, aural, input/haptic and other systems - are discussed in the next section, but a brief description of each type is given as follows:

2.4.1.1 Desktop VR

Desktop VR tends to be the cheapest form of VR system (see Figure 2.6. overleaf) as it consists mainly of a standard affordable computer which most people can now buy for home use. For this reason it is referred to as a ‘low-end’ system. The minimum specification of the computer in order to run the specialised VR software available (see section 2.4.2.2) is constantly changing.

Figure 2.6: An example of a Desktop VR system.

Three years ago it was possible to run a VE on a 33 MHz 386 PC with 8 MB RAM with a special graphics accelerator card and sound card. However, today, it is not worth purchasing a 386 PC even for the home market, as it would be difficult to maintain the parts, as many are no longer available and current software, not only VR software, would not be able to run effectively on it. Most computer vendors are now offering the home market, as standard, a Pentium 233 PC with 32 MB RAM and full multi-media capability, at an ever-decreasing price, making desktop VR more accessible to the home market, as well as industry. The usual visual display used by desktop VR is the computer monitor, generally SVGA (Super Video Graphics Adaptor) with anything up to 16 million colours and either 14”, 15”, 17”, 19” or 21 inches in size. However, desktop VR can support other visual devices like ‘shutter glasses’ and ‘stereo screens’ (see section 2.4.2.5). The sound cards in the computer are usually adequate for providing sound in the VE and are delivered through speakers. The input devices used to interact with the VE and often associated with this system are the keyboard, mouse, joystick or spacemouse (a six-degrees-of-freedom device - see section 2.4.2.7) although in some systems touch-screens are also used.

The main disadvantage of desktop VR is that the relatively small amount of processing power (compared with the more powerful computers used in other VR systems) places limits on the capabilities of the VE (this is discussed in section 2.4.2.1). Also, as the visual display is often just the computer monitor, these systems have been criticised for not utilising the full potential of the three-dimensional and 'presence' qualities of VEs. This is because the images are still essentially two-dimensional (unless the user has shutter glasses or stereo screens which can assist them to have stereo vision, discussed in section 2.4.2.5). Also, the VE does not fill the user's complete field-of-view and, therefore, it is still possible to get distracted by objects in the peripheral view which can diminish feelings of presence. However, this is still the most popular choice of system because the initial investment cost is minimal. Furthermore there are applications being developed using desktop VR to show that with careful consideration of the capabilities and limitations, effective VEs are possible. This is also a popular system because it has less of the possible side-effects which are associated with some of the other VR systems. Applications being explored are wide and varied, therefore in the short-term at least, desktop VR appears to be the way forward for many VE applications.

2.4.1.2 Projected VR

Projected VR systems use large projection displays, either on one surface (which can have a viewing angle of up to 120º across), or multiple projection displays to create a room or 'CAVE' (Cave Automatic Virtual Environment, see Figure 2.7).

Figure 2.7: An example of Projected VR (courtesy of the Centre for Industrial and Medical Informatics (CIMI), Nottingham, UK)

They run on high-end graphics workstations such as the Silicon Graphics range. The user can have shutter glasses in order to see the projections in three-dimensions and either some form of handheld device or a chair with hand devices attached are used to navigate around and interact with the VE. The advantage of these systems is that they increase the quality of immersion and presence, as compared with just a computer screen because the display fills the user's entire field-of-view, providing the illusion that they are 'in' the environment. However even though many people can see the VE, still only one user can control the navigation. Also they can be very costly in terms of equipment and development time and they require a lot of space. The types of applications currently being explored using this type of system have mostly been impressive walkthroughs of buildings, art galleries and museums (e.g. Shaw, 1994).

2.4.1.3 Artificial Reality or Video Mapping

An artificial reality system has been defined by Myron Krueger, as a system which "*perceives a participant's action in terms of the body's relationship to a graphic* *world and generates responses that maintain the illusion that his actions are taking place within that world*." (Krueger, 1991; p. 268). Krueger is recognised as the pioneer of this particular technology and there is no other researcher who has carried out more work into its applications. Comprehensive information about this system and its applications can be found in 'Artificial Reality' (Krueger, 1983) and 'Artificial Reality II' (Krueger, 1991). Generally, the system consists of videos and computers which merge a video image of the user with computer graphics. The user is able to watch a monitor which shows their body interacting with the objects (or people) in the VE (see Figure 2.8). This technology is already frequently used by the television media to provide interesting backdrops for the news, weather and other shows. It is also used by the film industry to merge computer-animated characters with real characters.

2.4.1.4 Augmented Reality

Augmented reality systems (see Figure 2.9) have been described, by Pimental and Teixeira (1992) as the "*use of transparent glasses onto which data, diagrams, animation or video can be projected to aid people who need to be simultaneously in the real world and also be able to access additional data to do their jobs" (p. 11).* The example application given by Pimental and Teixeira (1992) is of Boeing's exploration into the technology for aircraft engine mechanics. They are developing a system which allows the mechanics to access diagrams, parts lists and text while they work on a real engine, with the aim to be able to eventually overlay an entire structural diagram onto the engine to give the mechanic a type of 'x-ray' vision. Other application areas actively being explored are in the medical field by surgeons, using information from CAT (Computer-aided tomography) scans and x-rays to overlay onto real patients in order to examine the most appropriate route for surgery (Truppe et al, 1996).

2.4.1.5 Headset Systems

Finally, headset systems are probably those most associated with VR technology (see Figure 2.10, overleaf). They consist mainly of a head-mounted display/device (HMD) which the user wears to receive visual and auditory information from the

VE. The user interacts with the VE using some form of hand-held controller, like a wand, joystick or dataglove, and trackers on the HMD and input devices allow the computer to constantly update the position of the user. The advantage of this type of system is that it completely blocks any external influences from the real world by enclosing the user's visual and auditory senses with a HMD. This creates a strong sense of immersion and presence, as the user is unaware of any other environment except the VE. However, for this very reason, the system also has its disadvantages. Researchers in the area believe that the closer that the VE gets to resembling reality the more cases of 'simulator-type sickness' may occur (Kennedy, 1991; Levison and Pew, 1993).

Figure 2.8: An example of an Artificial Reality System (courtesy of Myron Kreuger, Artificial Reality, US)

Figure 2.9: An example of Augmented Reality (courtesy of MIT AI Lab and Surgical Planning Lab, Brigham and Women's Hospital, US).

Figure 2.10: An example of Headset system, the Elysium system developed by Virtuality, UK.

Simulator sickness has been defined as *"a feeling of discomfort that arises from performing tasks in the simulator, where such discomfort is not elicited when the same tasks are performed operationally. This discomfort may include nausea and disorientation that occur while the simulated tasks are being performed, plus adverse symptoms that persist (or become initially apparent) after the person has left the simulator."* (Levison and Pew, 1993; p. 70). It is also sometimes referred to as 'cybersickness' and is believed to be partly caused by 'sensory cue conflict', that is, when a person's senses are receiving conflicting information. In a VE, the user may be 'moving' and receiving all the usual information via their visual and auditory senses to confirm this. However, in the real environment they are stationary and their body systems are sending contradictory information. It is felt that this conflict is enough to induce ‘sickness’.

However, much research is on-going in this area in order to identify the factors which may cause this effect and ways of predicting an individuals’ susceptibility to simulator sickness (Hettinger et al, 1990; Kennedy et al, 1992; McCauley and Sharkey, 1992; Kennedy et al, 1993; Oman, 1993; Regan and Ramsey, 1994a; Regan and Ramsey, 1994b; Cobb et al, 1995). Therefore, this type of system has a number of problems which need to be addressed before it should be widely implemented.

2.4.2 Components of VR systems

Given the types of VR systems listed before, generally they consist of a hardware platform and software application linked via tracking systems to visual, auditory, haptic and/or other systems that allow the user or users to interact with the VE and to receive information from the VE. Some of the main technologies used to form the different VR systems are discussed briefly below.

2.4.2.1 Hardware Platforms

Hardware platforms are the computer systems which support the specialised VR software and the peripherals (i.e. the visual, audio, haptic and any other systems). There is a wide variety of hardware platforms that VEs can run on and the cost of these platforms are constantly decreasing making the entry level to VEs more accessible to more companies. At the ‘low-end’, VR systems can use a standard PC, which in today’s terms (as stated before) - 1998, refers to a Pentium 233 PC with 32 MB RAM, SVGA monitor, full multimedia capability (i.e. can display photo-realistic graphics and video with good quality stereo sound), costing around £1500 and decreasing all the time. There is also some VR software available for Apple Macintosh systems. At the high-end, VR systems use some of the leading graphic workstations or UNIX systems, e.g. Silicon Graphics’ ‘RealityEngine2’, the University of North Carolina’s ‘Pixel Planes 5’ and ‘PixelFlow’ and the Evans

& Sutherland's 'Freedom Systems'. These systems run on parallel architectures, that is, they can process graphics on parallel paths so that they can generate complex images at high speeds. These systems can cost hundreds of thousands of pounds but again the prices are also falling rapidly. Therefore ironically unlike some other applications which use computers, the cost of the computer hardware for a VR system is relatively low compared to the other parts of the system.

The choice of hardware platform should be based on the application requirements as this will determine the design of the VE and what peripherals are necessary for the user to be able to fulfil these requirements. In particular, the appropriate hardware platform has to provide the necessary 'frame rate', 'response time' and 'resolution' (NRC, 1995), in order to maintain the real-time, interactive, immersive and presence qualities of the VE for the user. 'Frame rate' is defined as "*the number of still images that must be presented per second to provide the illusion of continuous motion*" and 'response time' (also referred to as lag, latency, or total delay), is the ability of the VE to provide an instantaneous response to an action (NRC, 1995; p. 57). The National Research Council have summarised current research to conclude that "*Frame rates must be greater than 8 to 10 frames/second. Total delay must be less than 0.1 second*." (NRC, 1995; p.250) to maintain the quality of the experience. Finally the resolution of an image, that is, how much detail the image has, needs careful consideration. According to Bricken, M. (1991), the human eye views a real world image at about 80 million polygons (a flat plane figure with multiple sides used as the building blocks of VEs) at over 30 frames a second. However, systems today can just about generate ten thousand polygons and any more would slow them down considerably. Therefore, the application requirements and the subsequent design of the VE affects the hardware platform chosen.

2.4.2.2 VR Software

VR software are specialised computer programs which allow a VE to be built and used. There are two main types of VR software available - toolkits and authoring systems. Toolkits are programming libraries, generally for C or C++, which provide a set of functions that a VE developer can use to develop VEs. These programming libraries are generally more flexible and provide faster rendering than authoring systems, however they require a highly skilled computer programmer. Authoring systems are complete programs which use a basic ‘scripting’ language which describe complex actions which can be directly assigned to the appropriate objects. They require less detailed programming and therefore are much easier to learn. The main types of VR software available (compiled by Feldberg, 1997) are listed in Table 2.2.

Table 2.2 : Currently Available Virtual Reality Software (Feldberg, 1997)

Product	**Company**	**Platform**	**Cost**
ACK3D	Written by Lary Meyer	DOS	Freeware
Gossamer	Written by Jon Blossom	Macintosh	Freeware
MRToolkit	Written by University of Alberta	UNIX	Freeware
Multiverse	Written by Robert Grant	UNIX	Freeware
Rend386	Written by Dave Stampe & Bernie Roehl	DOS	Freeware
VEOS	Written by HITLab	UNIX	Freeware
2Morrow Tools	2Morrow	DOS	$100
Virtual Reality Studio (VRS)	Domark	DOS	$100
Lepton VR Data Modeling Toolkit.	Lepton	DOS	$150
Qd3d, 3dPane, & SmartPane C++ libraries	ViviStar Consulting	Macintosh	$192
MindRender (API)	ThemeKit Ltd.	DOS	$150 - $600
VREAM	VREAM	DOS	$595
WorldToolKit	Sense8	SGI, Sun, E7S.	$795 -

		DOS. Windows, NT	$12,500
Virtus Walkthrough	Virtus Corp	Macintosh	under $1000
Lightscape	Lightscape Graphics Software	SGI	over $1000
PhotoVR	Straylight Corp.	SGI	over $1000
Superscape VRT	Superscape	DOS, Windows '95	$3995
dVise	Division	SGI	$55,000 - $200,000

The following information summarises the descriptions on the software provided by the internet sites of the HITLab, Chris Hand and Ian Feldberg all up-dated at the end of 1997.

The information in Table 2.2 is listed in terms of price. The price scale begins with 'freeware' software. This is software which can be downloaded from various internet sites, free of charge, but the original programmers retain copyright, which means that commercial use is restricted. The main freeware software includes 'ACK3D' and 'REND386' for the standard PC; 'MRToolkit', 'Multiverse' and 'VEOS' for UNIX systems and 'Gossamer' for the Macintosh. They tend to be quite basic (most begin as student projects) but they are widely known in the VR community. In particular, REND386 appears in 'The Virtual Reality Casebook' (Loeffler and Anderson, 1994). Also they allow people to explore developing VEs without having to initially invest a large amount of money.

The next price level consists of software which is commercially available for under two hundred dollars. This includes 2Morrow Tools, VRS or 3D Construction Kit, Lepton VR Data Modelling Kit and MindRender for PC systems and Qd3d, 3dPane and SmartPane for the Macintosh. In general, these have been used to develop 3D computer games to provide high quality graphics but they do not have much scope for complex interactions.

However, the software which costs between two hundred and one thousand dollars, such as VREAM, WorldToolkit and Virtual Walkthrough, do offer complex interactions. Also, they generally do not require any specialised hardware beyond a basic computer and can support many different input devices. In particular, WorldToolKit is currently one of the most popular choices of VR software, (especially for programmers) even though it requires high level computer skills.

Finally, the professional VR software packages begin at about one thousand dollars upwards. The hardware required to run these packages varies from standard computers to workstations. Lightscape and PhotoVR are software packages generally aimed at architects and designers who want to create 'photo-realistic' walkthrough buildings. Superscape and dVise are probably the best known in the VR market. Next to WorldToolKit, Superscape VRT is probably the most widely used authoring system because it does not require a high level of programming skill or dedicated systems. However, dVISE (which is similar to VREAM) does require a high level of programming skill and expensive hardware.

In general the software listed in Table 2.2 all have varying advantages and disadvantages. Apart from cost and hardware already available, choice is again based on the requirements of the application, and to a lesser extent, the capabilities and limitations of the VE developer. The different packages have different methods of programming, produce different kinds of VEs (mostly visually) and support different peripherals. Therefore, a careful decision must be made, as it may limit what is required for successfully fulfilling the application.

2.4.2.3 Tracking Systems

Tracking systems are position and orientation sensors which link the user's position in the real world with their 'representation' in the VE. They are required to provide three measures of position (x, y, z) and three measures of orientation (roll, pitch, yaw). Pimental and Teixeira (1992), list several key parameters which determine their effectiveness:

- *lag or latency* (also referred to as response time or total delay) which is the delay between sensor measurement, processing and delivery to the computer. As discussed before (section 2.4.2.1), it is suggested that total delay should be less than 0.1 second (NRC, 1995) otherwise performance is effected;
- *up-date rate,* which is the speed at which measurements are made and it is not directly related to lag. For example, the system may have a lag of 100 msec but still send 100 measurements a second back to the computer;
- *interference,* which is the sensitivity of the tracker to environmental factors, such as sound, temperature etc.;
- *accuracy,* which is the reliability of the position and orientation information given by 'translational' values, which vary by inch, and 'rotational' values, which vary by degrees; and
- *range,* which is the maximum distance between the user (or object) and sensor to maintain a specified accuracy.

The main position and orientation trackers available for VR systems (compiled by Feldberg, 1997) are listed in Table 2.3.

Table 2.3: Currently Available Position and Orientation Trackers (Feldberg, 1997)

Product	Company	Type	Freq	df	Lat.	Rn.	Acc.	Price
InsideTrak	Polhemus	Mag.	30Hz	6	12ms	5ft	0.5in 2.0º	$999
InsoTrak II	Polhemus	Mag.	30Hz	6	20ms	5ft	0.1in 0.75º	$2875
FasTrak	Polhemus	Mag.	30Hz	6	4ms	10ft	0.03in	$6050

							0.15º	
Flock of Birds	Ascension	Mag.	144 Hz	6	n/a	3ft	0.1in 0.5º	$2695
Flock of Birds/10	Ascension	Mag.	144 Hz	6	n/a	10ft	0.1in 0.5º	$8090
VR-360	Angularis	Inert.	500 Hz	3	2ms	20ft	n/a	$9200
V-scope	Eshed Science & Tech.	Ultra	100 Hz	3	2ms	12ft	n/a	$2800
Cyber Track	General Reality	Inert.	30Hz	3	<50ms	n/a	1.25º	$???

(Mag. = Magnetic; Inert. = Inertial; Ultra = Ultrasonic; Freq. = Frequency; df = degrees of freedom; Lat. = Latency; Rn. = Range; Acc. = Accuracy)

There are four basic types of trackers - magnetic, ultrasonic or acoustic, mechanical and optical systems. These are often used in combination in most VR systems because they have different strengths and weaknesses. Detailed information about these systems can be found in Aukstakalnis and Blatner, (1992); Boman et al, (1992); Pimental and Teixeira, (1992); Meyer et al, (1992); Burdea and Coiffett, (1994) and the NRC, (1995) and is summarised as follows.

Magnetic or electromagnetic trackers are the most popular method of tracking because of their small size and freedom of movement. They use a set of three wire coils which emit a high frequency to produce magnetic fields. When a similar set of coils is placed in the range of this field a small voltage is induced and this is translated into a measurement of position and orientation. This technique was first developed by Polhemus for military applications and is often used in a headsets. The main limitations are their moderate accuracy, short range, the delay between measuring and processing information (or high latency) and their susceptibility to magnetic interference. However, they are still convenient and have a low cost.

Ultrasonic or Acoustic sensors use three ultrasonic transducers and three small microphones. Transducers convert an action into a form which can be interpreted by a computer. The transducers are placed at three points about a foot apart to form a triangle. The microphones form a smaller triangle and become the sensor which is usually placed on the top of a headset. The ultrasonics emit a high frequency which is sensed by the microphones and their relative position is recorded and processed. Again, there are limitations on accuracy and they have problems with interference from echoes and other noises in the environment, but they are inexpensive and have minimal latency.

Mechanical tracking systems use a direct mechanical connection between a reference point and the user (or object) being tracked. The link is usually a mechanical arm with rotating joints, allowing six degrees-of-freedom, or they may be highly complex exoskeletons (for more detailed positions). Their major disadvantages are the 'encumbrance' of the equipment which links the user with the reference point, and it also restricts motion. However, these systems are relatively inexpensive, provide reasonably accurate tracking and have very small latency.

Finally, optical position tracking systems use infra-red technology and are currently not widely used. There are a number of different methods being explored. One such method being developed at the University of North Carolina (Pimental and Teixeira, 1992), involves a ceiling grid of LEDs (Light-emitting diodes) and a head-mounted camera. The LEDs are 'pulsed' in sequence and the camera's image is processed to detect the flashes. Two problems with this method are the limited space (grid size) and lack of full motion (rotations). Another method uses a number of video cameras to capture simultaneous images which are then correlated by high speed computers to track objects. The major limitations of these systems are their need for high-end computers for the necessary processing power and also the visibility constraints. However they are extremely fast and accurate.

A further category of tracker, which is rarely considered for VR systems, is the inertial tracker which has been used for long-range navigation for years. Two miniature gyroscopes are mounted on a HMD and as the user moves, the inertial weights in the gyroscope's spinning motor reflect this via an LED (Light Emitting Diode). An optical sensor picks up the pattern and translates it into positions. They generally only provide rotational measurements but they are very accurate, unconstrained by range, interference and also, latency is low.

2.4.2.4 Visual Systems

The visual displays used in VR systems have already been briefly mentioned in the discussion of the types of systems available (section 2.4.1). At the low-end of the market, there is the standard computer monitor used in desktop VR which generally comes in five sizes 14", 15", 17", 19" and 21 inches. Projected VR and artificial reality tend to use video projection screens, blue screens or white surfaces of various sizes (up to room-size in the case of CAVE systems). All these systems use visual displays which have been developed for many years now. However, augmented reality and headset systems use display devices which are still in the development process and, therefore, have problems which need addressing. These are discussed later on.

The development of headset systems has been the focus of much attention since Ivan Sutherland's work in the late 1960's. Since then, research and development has been carried out in many places in the US mainly the Massachusett's Institute of Technology (MIT), NASA/Ames and Langley Research Centres, Wright-Patterson Air Force Base, the Naval Ocean Systems Centre, the University of North Carolina, LEEP Optics, the University of Washington, CAE Electronics, VPL Research, Virtual Research, Technology Innovation Group, Kaiser Electronics Electro-Optics Division, Hughes Electro-Optical and Data Systems Group, Stereographics Corporation and Fake Space Labs (NRC, 1995) and also in Japan at the government's Mechanical Engineering Laboratory at Tsukuba. The result of this work has led to many head mounted devices (HMDs) and off-head displays (OHD) now widely available, as shown in Table 2.4 overleaf (compiled by Feldberg, 1997). As is shown, the main considerations when selecting a visual device (excluding cost, as this has the obvious effects on selection) are the maximum resolution available, the field-of-view (FOV) and the technology of the system.

Table 2.4: Currently Available Visual Displays (Feldberg, 1997)

Product	**Company**	**Res.**	**FOV**	**Type**	**Price**
FOHMD (HMD)	CAE-Electronics	1000x1000	127H x 66V	fibre-optic	$250,000
dVISOR (HMD)	Division	345x259	105H x 41V	colour LCD	$5000
BOOM-2C (OHD)	Fake Space Labs	1280x1024	90 - 100H. 100% overlap	dual colour CRTs	$74,000
BOOM-3C (OHD)	Fake Space Labs	1280x1024	90 - 100H. 100% overlap	colour CRTs	$74,000
VFX1 (HMD)	Forte Technologies	428x224	46H x 35V	colour LCD	$695
CyberEye (HMD)	IMPART	400x300	22.5H x 17V	colour LCD	$1995
Sense Cover cx2 (HMD)	Kaiser Electro-optics	1280x1024	120H x 90V, 60% overlap	colour LCD	$8700
Sim Eye 60 (HMD)	Kaiser Electro-optics	2340x230	60diag up to 60 x 100	colour LCD	$135,000
Vim 1000pv (HMD)	Kaiser Electro-optics	2340x230	100H x 30V, 100% overlap	colour LCD	$6,495
Vim 500pv (HMD)	Kaiser Electro-Optics	780x230	40H x 30V	colour LCD	$2,495

Cyberface 2 (OHD)	LEEP	479x234	140H 100% overlap	colour LCD	$8100
Cyberface3 (OHD)	LEEP	720x240	70H	colour LCD	$15,000
MRG 2.2 (OHD)	Liquid Image	240x240	84Hx65V	colour LCD	$3,495
MRG 4 (OHD)	Liquid Image	480x234	61Hx46V	colour LCD	$2,195
MRG 3c (OHD)	Liquid Image	768x556	84Hx65V	colour LCD	$5,500
MRG 6 (OHD)	Liquid Image	640x480	40º diag	grey scale	$3495
Datavisor 80 (HMD)	nVision	up to 1280 x 1024	80º	colour CRT	$100,000+
Datavisor HiRes (HMD)	nVision	up to 1280 x 1024	~ 52º	colour CRT	$24900
Datavisor VGA (HMD)	nVision	640x480	~ 52º	colour CRT	$24900
Virtual Binoculars (OHD)	nVision	up to 1280 x 1024	60º	colour CRT	???
Private Eye (OHD)	Reflection Technology	720x280	22H x 14V	mono LED	$500
HMSI Model 1000 (HMD)	RPI	450x240	45H x 32V 100% overlap	colour LCD	$5000

(HMD = head mounted display/device; OHD - off-head device/display; Res. = resolution; FOV = field-of-view)

Table 2.4 (continued): Currently Available Visual Displays (Feldberg, 1997)

Product	**Company**	**Res.**	**FOV**	**Type**	**Price**
CrystalEyes (OHD)	Stereo Graphics	n/a	n/a	LCD shutter glasses	$985
SGS (OHD)	Tektronix	n/a	n/a	passive glasses	$1500
CyberMaxx 2.0 (HMD)	VictorMaxx **discontinued**	780 x 230 180Kp	56º	colour LCD	$899
i-glasses Pro (HMD)	Virtual I/O **discontinued**	180Kp	30º	colour LCD	$799- w.tracking $599 - video
i-glasses Home (HMD)	Virtual I/O **discontinued**	113Kp	23º	colour LCD	$599- w.tracking $399 - video
PID P1 (HMD)	Virtual Reality	479x234	63H, 100% overlap	colour LCD	$5000
PID 131 (HMD)	Virtual Reality	1280x1024	40H x 30V 100% overlap	mono CRT	$56000
PID 133 (HMD)	Virtual Reality	1280x1024	40H x 30V 100% overlap	colour CRT	$87000
VR4 (HMD)	Virtual Research	742x230 170Kp	60º	colour LCD	$79000
FS5 (HMD)	Virtual Research	800x600	55º	colour CRT	$19,900
RSP110 (OHD)	Virtual Vision	300x200	20H	colour LCD	$700
DK210 (OHD)	Virtual Vision	300x200	60H, 100% overlap	colour LCD	$2900

(HMD = head mounted display/device; OHD - off-head device/display;

Res. = resolution; FOV = field-of-view)

As discussed in section 2.4.2.1, 'resolution' is concerned with the detail in an image. In terms of how this detail is displayed on the screen, the measurement is given in pixels (defined as, "*a contraction of picture element, it refers to one point in a graphics image on a computer display*" Pimental and Teixeira, 1992). The more pixels a display has, the more detail it can portray. A standard SVGA computer monitor has 800 x 600 pixels but as shown in Table 2.14, there are many systems which offer a much higher resolution (these systems are usually also the most costly). However, the greater the resolution, the more processing power is required, which can cause lag and therefore effect the user's performance. Therefore, it is sometimes not advisable (or necessary) to use the highest resolutions, but usually the system will allow the VE developer some control in setting the appropriate level.

Also important is the field-of-view (FOV), offered by the system. FOV is the range the eye can see and not only what the eye can focus on. Both eyes provide a lateral FOV of approximately 180º horizontally (with some overlap between them) although, because we are able to 'pivot our eyes' a bit further, we can actually perceive approximately 270º (Aukstakalnis and Blatner, 1992). The FOV offered by VR display systems are measured horizontally and vertically or by degrees. The appropriate FOV is dependent on the application, for example, desktop VR offers a very narrow FOV, so it is unsuitable for applications where it is necessary to detect objects in the peripheral view or for applications where it is necessary that the user does not get distracted by external influences.

Finally the technologies used for the displays are most frequently 'cathode ray tubes' (CRTs) and 'liquid crystal displays' (LCDs). CRTs are common imaging devices found in most television and computer screens. They produce their own light activated by an electron beam and are able to provide extremely high image resolution. Another advantage is a CRT's brightness which makes it ideal for using in augmented reality applications like in head-up displays (HUDs) used by the military for pilots (Pimental and Teixeira, 1992). However, their major disadvantage is that they are monochrome. Some research has been carried out to look at using colour lenses, but it requires the system to generate 180 frames per second - three times faster than a standard display (Pimental and Teixeira, 1992). Also they require high voltages to be used which are positioned closely around the user's head. LCDs however are relatively more compact and inexpensive to produce. They also generate good colour at low level currents. They are 'light modifiers' rather than producers, so when there is no electrical current, no light is allowed to pass through them and the pixels remain dark. To create colour, light filters of usually red, green and blue are used. If light is allowed through a pixel it then passes through a colour filter and the combination of red-, green- and blue-lit filters creates the illusion of multiple colours. LCDs are commonly used in watches and calculators. However, they have very limited resolution and if the FOV is widened the viewpoint degrades, therefore a trade-off between high resolution and FOV has to be considered. Both CRTs and LCDs, though, still produce quite bulky headsets because of the limited technology available for producing small versions of the screens.

An alternative approach is to use fibre-optic cables which can produce very lightweight HMDs and have been used in the medical field for endoscopes (a device consisting of a tube and optical system used to explore hollow organs and cavities). A high resolution colour image is projected at one end of a fibre-optic cable (approximately one inch thick and containing at least a million fibres) and this image is directly viewed by the user. The major disadvantage is the high cost of producing such a system and price is dependent on the amount of fibres used. Therefore the current technologies available place constraints on visual systems in terms of quality of resolution, colour of images, safety and weight of the headset.

Furthermore, the visual displays can provide either monoscopic or stereoscopic vision. Monoscopic vision is where the user views the same image of the VE with both eyes. Even though in essence the images are two-dimensional, it is still possible to experience the VE as three-dimensional, through several depth cues. These are listed by Aukstakalnis and Blatner (1992) and Friedhoff (1989) cited by Pimental and Teixeira (1992), as follows:

(1) Static Depth Cues

- *interposition or occlusion* - objects in the foreground 'occlude' or are positioned in front of objects in the background;
- *shading* - a shaded edge provides information about shape, as well as, volume;
- *brightness or aerial perspective* - objects at a distance appear duller or 'bluer' than objects close by;
- *size* - objects which are closer appear bigger than objects which are at a distance;
- *linear perspective* - for example, the sides of a building appear to converge at a distance; and
- *texture gradient or detail perspective* - objects become less detailed the more distant they are.

(2) Motion Depth Cues

- *motion parallax* - as you move, objects closer move faster than objects at a distance

(3) Physiological Depth Cues or Occulomotor Cues

- *accommodation* - this refers to the flexing of muscles which change the shape of the lens in the eye. The muscles are quite relaxed when viewing distant objects, however, when focusing on closer objects the muscles are more active; and
- *convergence* - this refers to how far the eyes must turn inwards to view an object.

However, our eyes more readily perceive depth by stereoscopic cues through a process called 'stereopsis' or 'binocular disparity', (this was proved by research carried out by Bela Julesz (1971) cited by Pimental and Teixeira, 1992). This is where different images of the world are received by each eye, the same environment, but slightly set apart. How the brain then merges these images is still uncertain, however it allows us to perceive the environment in three-dimensions. In the same way VR display systems offer stereoscopic vision by generating two different images of the VE and off-setting them by the equivalent distance between the eyes. In the headset, the two different images are directly delivered to the appropriate eyes by providing two sets of screens. For desktop VR and projected VR, the images are placed side-by-side and the user can be 'assisted' to cross their eyes. This can be done by providing the user with polarised glasses (equivalent to red/blue glasses) and projecting the images through polarized filters. Alternatively the images can be rapidly displayed sequentially and 'shutter glasses' can be used to 'close' alternate eyes so that the images appear to be fused. Other methods involve split screen techniques, where the monitor is either divided into two parts and a special hood viewer is placed against the monitor to position the eyes correctly or the images are oriented so that the top of each points out the sides of the monitor and a special hood containing mirrors is used to correctly orient the images.

In general the current 'off-head displays' (OHD) used for VR systems, (these include, monitors and projection screens) are generally inexpensive and well-developed and have limited problems associated with them. However headset displays appear to be expensive and have a number of issues which need to be addressed. These were identified by the National Research Council (NRC, 1995) as:

"1) high-resolution, miniature, lightweight, lowcost display surfaces are yet to be realised.

(2) weight and inertial burdens imposed by most HMDs affect the incidence of symptomatic motion sickness, the ability of users to make proper judgements concerning orientation and their long-term habitability;

(3) due to size, performance and cost constraints, fixation/focus compensation is utilised in most HMDs and conflicting visual depth cues are provided to the user. Furthermore the proper operation of HMDs is intimately tied to the performance of headtracking systems (i.e. update rate and lag) which is currently less than ideal." (p. 131)

2.4.2.5 Audio Systems

The technologies for audio systems - earphones, headphones and external speakers - have been developing over many years for the audio industry, as well as for use by the home market. The most popular method of controlling and generating sounds uses the MIDI (musical instrument digital interface) standard. Sounds are first digitally sampled or converted from analog to digital form then played back using a sequencer. A variety of sounds can be sampled - voices, musical instruments and special effects - and further modifications can be made e.g. changes in pitch or the sound envelope or reverberations (echoes). (Pimental and Teixeira, 1992). However there are significant problems when matching sound to visual information as required in a VE. Firstly, every object and event must have its appropriate sound which requires complex programming and a huge amount of storage space for the data (NCR, 1995). This can cause lag in the system and increases development time of the VE. Secondly, as visual frame-rates often change in line with the viewpoint of the user (i.e. if the view has little complexity then the frame-rates are faster and the alternative is also true), the sound must also keep in line with these changes, but play back of recorded sounds at different rates often leads to distortions (Pimental and Teixeira, 1992). Further difficulties lie with producing three dimensional or localised sound. There are several factors which control a person's ability to localise sound. These include (from the work performed by Bergault, 1987; Greuel, 1991; cited by Pimental and Teixeira, 1992):

- *Interaural time difference* - this refers to the time difference between each ear receiving the same sound;
- *Interaural amplitude difference* - this refers to the difference in sound pressure or loudness received by each ear;
- *Frequency difference or acoustic shadowing* - this refers to the fact that higher frequencies can be easily blocked then lower frequencies; and
- *Head-Related Transfer Functions* (HRTF) - this refers to the 'convolutions' of the pinnae or outer ear which gather sound and reinforce certain sound frequencies.

Other factors include (Aukstakalnis and Blatner, 1992):

- *visual aids* - this refers to the fact that our vision provides us with a good idea of where sound may be coming from;
- *sound familiarity* - this refers to our ability to remember certain sounds; and
- *echolation* - this refers to environmental factors which echo or reverberate to provide us with information e.g. a small room has less echoes than a huge empty hall.

The main problem in VR systems is convincing the user that the sound is coming from the 'virtual object' rather than the earphones in their headset or the speakers. The audio industry for many years have been developing external speakers in order to produce a surround effect although this can only occur in a limited area. Most of the above factors though can be reproduced to a certain extent but the most difficult is the HRTF. There are however some systems available that attempt this problem. These are listed in Table 2.5 overleaf (compiled by Feldberg, 1997). As Table 2.5 shows the market is mostly dominated by the work carried out by Crystal Rivers Engineering, in particular, Scott Foster and Elizabeth Wenzel in collaboration with NASA's Ames Research Centre (Foster et al, 1991; Wenzel, 1992). This work has been on-going since 1985, but these systems still have their limitations, mainly due to the limits of processing power available.

Table 2.5 Currently Available Sound Generators (Feldberg, 1997)

Product	**Company**	**Input**	**Price**
Convolvotron	Crystal River Engineering	4 channels	$15,000
Beachtron	Crystal River Engineering	4 channels	$1495
Alphatron	Crystal River Engineering	2 channels	$495
Acoustetron II	Crystal River Engineering	8 channels	$11,995
Focal Point	Focal Point 3D Audio	2 channels	$1500
Audio Architect	Visual Synthesis	2 channels	$500
Audio Image Sonic Architect	Visual Synthesis	2 channels	$1500
Audio Image Sound Cube	Visual Synthesis	2 - 8 channels	$8000

Other research in the area of audio systems has also looked at the use of sounds as metaphors for other sensations (Massimino and Sheridan, 1993) e.g. a beep if something has a hot temperature or is dangerous. Further research is required to develop this as currently it is difficult to reproduce 'feel' in VEs

2.4.2.6 Haptic and Input Systems

Haptic systems refer to *"all the physical sensors that provide us with a sense of touch at the skin level and force feedback information from our muscle and joints"* (Pimental and Teixeira, 1992). Input devices, also known as interaction devices, are those that allow the user to "*navigate, select, interact and command*" in a VE (Pimental and Teixeira, 1992). The simplest input devices are: the keyboard, conventional mouse, a spacemouse, trackball and a joystick. These are the devices often used with desktop VR and can be seen in Figure 2.6. In terms of tactile and force feedback technologies, the systems currently available are shown in Tables 2.6 and 2.7 overleaf (compiled by Feldberg, 1997). In general there is a limited amount of commercially available tactile and haptic systems, due to the difficulties in developing such devices and our limited knowledge of our own body system. (Details of each device can be found in Hand, 1997 and links with the main developers in the HITLab, 1997).

Table 2.6 Currently Available Haptic Devices (Feldberg 1997)

Product	**Company**	**Feedback**	**Body Part**	**Price**
Teletact II	ARRC/Airmuscle	Force	fingers, hand	$4900
Interactor	Aura Systems	Tactile	chest	$89
DTSS X/10	CM Research	Tactile	skin	$10,000
TouchMaster	Exos	Tactile	fingers	?
Force ArmMaster	Exos	Force	shoulder, arm	?
SAFiRE	Exos	Force	fingers, wrist	?
Exoskeletal Master	Sarcos	Force	fingers, hand	$100,000
Phantom	SensAble Technologies	Force	one finger	$19,000
CyberTouch	Virtual Technologies	Tactile	hand (fingertips ,palm)	$14,800
Tactools XTT1	Xtensory	Tactile	skin	$1500

Table 2.7 Currently Available Glove Techologies (Feldberg, 1997)

Product	**Company**	**Type**	**Sensors**	**Price**
Dexterous Hand Master	Exos	Mechanical	20	$15,000
5th Glove	Fifth Dimension Technologies	Fiberoptic	5	$495
Exoskeletal Hand Master	Sarcos	Mechanical	16	?
CyberGlove	Virtual Technologies	Resistive bend sensing	18 or 22	$9,800 or $14,500
Dataglove	Greenleaf Medical Systems (formerly VPL)	Fiberoptic	10	?

Some of the problems are associated with accurately relating actions in the real world with actions in the VE causing delays to response and impairing the experience. Also the devices can be quite difficult to wear which can further effect performance. Finally, the feedback received tends to involve just pressure however touch involves other sensations like temperature and 'wetness', so this feedback is not complete. However the ability to 'virtually feel' an object in an environment can promote the feelings of immersion and presence and highlight realism of an action. Also it may be necessary in some applications therefore it is an area which requires attention.

2.4.2.7 Other systems

Research work is also being carried out in the area of real world motion including: (1) whole-body passive motion (passive transport); (2) whole body active motion (locomotion); and (3) part-body active and passive motion (e.g. when an arm is moved passively or actively) (NCR, 1995). Passive systems can be divided into two categories - 'inertial displays' where the user is actually moved e.g. fair ground rides; and 'non-inertial' displays in which motion is simulated without the user moving e.g. flight simulators. Motion systems for active transport (i.e. locomotion) provide the user with the experience that they are perceiving in the VE e.g. the use of a treadmill to provide the sensation of walking or running through the VE.

However there is limited research by the VR community on any of the other senses like smell and taste. However these may be important for applications like fire-fighting where it is necessary to recognise gas leaks, etc. (NRC, 1995).

2.5 Discussion

This chapter has outlined the main definitions of Virtual Reality (VR) and Virtual Environments (VEs), presented a brief overview of how the technology has developed and then detailed the current status of the different components of the technology. Generally it is possible to see from this chapter, that there is a substantial gap between the definition of the capabilities of virtual environments (VEs) and the actual capabilities of the currently available technology.

VEs were defined at the beginning (section 2.2.1) as having the following attributes (based on Wilson et al, 1996):

- computer-generated representations of real or imaginary environments
- experienced as three-dimensional via a number of senses - visual, aural and/or tactile.
- objects within these environments are independent of the user and can display real world behaviour
- the user or users have autonomous control - the freedom to navigate and interact with objects, using a number of different viewpoints.
- this interaction occurs in real-time
- the users experience feelings of presence and/or immersion in the VE

Many of these attributes however are significantly restricted by the limitations of the technology. A summary of the main capabilities and limitations of the components of VR, discussed in this chapter, is shown in Table 2.7.

Table 2.8 Summary of the main capabilities and limitations of current Virtual Reality systems

	Capabilities	Limitations
Hardware	• a wide variety of platforms to choose from • price is ever-decreasing while capabilities (graphics, sound, videos) are ever-increasing • processing speeds are increasing	• choice of platform places a constraint on the types of components chosen for VR system • processing speeds still impose limitations on all other systems in terms of frame-rates, response times and resolution.

Software	• a wide variety of software to choose from catering for different levels of computer skill • price ranges from 'freeware' to costly but sophisticated software	• software does not integrate particularly well with other computer software like CAD • capabilities of each software can place constraints on the design of the VE

Table 2.8 (continued): Summary of the main capabilities and limitations of current Virtual Reality systems

	Capabilities	Limitations
Tracking systems	• choice of mainly four systems at varying costs and abilities	• problems with accuracy, range, response time and interference have to be considered for each type
Visual systems	• wide variety of systems to choose from • low cost options are well-developed (e.g. monitors, projection screens) • high-cost options provide stereo vision and promote immersion and presence (e.g. augmented reality displays, headsets)	• problems to consider include quality of resolution, field-of-view, stereo vs. mono. vision • particular problems with headsets include: cost, weight, side-effects
Auditory systems	• technology has already been developed for years • use of sound to replace another sense • increase realism of objects and situations	• costly in terms of computer power • difficulty in matching sound directly with visual events • difficulty in spatial location of sounds

	• increase immersion and presence	
Haptic and Input systems	• conventional input devices are well-developed	• limited choice of tactile and force feedback devices • still in development with many problems including: fit, accuracy, type of feedback
Other systems	• motion systems	• little work on smell and taste

What is immediately apparent is that each component of a VR system has its own capabilities and limitations, but it is not possible to consider each component in isolation. There is very much an interaction between all the different components which either enhances the strengths of the VR system or places further limitations on their weaknesses. For example, if a high quality stereoscopic headset is required, immediately this excludes any low-end computer system as they would be incapable of supporting such equipment. Some form of high-end workstation is required that is compatible with the headset and has enough processing power to generate two separate images simultaneously while maintaining resolution, response time and minimising lag. Therefore it is possible to see that significant improvements in all the components of a VR system would benefit greatly VE applications and the need for research into how the limitations of the VR system can effect performance. However, not all the components of a VR system need significant improvements. For example, some of the visual displays (monitors, projection screens, etc.), auditory displays (headphones, speakers, etc.) and input devices (keyboard, mouse, joystick, etc.) have been fully developed for other purposes for many years. As such some VE applications have been successfully developed and implemented, although generally only by the entertainment industry but it shows real possibilities for industry.

The next chapter deals with these possibilities by first presenting the needs and requirements of industry for VR and VEs, as expressed by industry themselves and then presenting the current application areas being researched by academic and industrial groups around the world.

Chapter 3: Current Application and Research Areas of Virtual Environments

3.1 Introduction

This chapter examines the current research and development areas of virtual environment (VE) application. It leads on from the overview of virtual reality (VR) technology (discussed in chapter two), to consider firstly in section 3.2, a re-assessment of the work carried out for the Engineering and Physical Sciences Research Council (EPSRC) which identified industry's understanding, application requirements, likely implementation and needs of VEs. This is followed by section 3.3 which presents a breakdown of the application areas which are currently being explored by research and development groups around the world in academic institutes and industry and the chapter concludes with a discussion of the main points in section 3.4.

3.2 Virtual Reality for Industrial Applications: Opportunities and Limitations

During 1994 - 1995, the Engineering and Physical Sciences Research Council (EPSRC) funded a study (awarded under grant GR/J57643) entitled: "Applications of Virtual Reality in UK Manufacturing Industry: A Feasibility Study". This was awarded to VIRART (Virtual Applications Research Team) in the Department of Manufacturing Engineering and Operations Management, at the University of Nottingham. (Subsequent reports from this work are listed under the references as Cobb et al 1994a, 1994b, 1994c, 1994d; and D'Cruz et al, 1994, Wilson et al, 1994, Wilson et al, 1995). This study was part of the MOVE (Manufacturing Operations in Virtual Environments) programme which was initiated by VIRART to identify the needs and address them of manufacturing industry. Although the remit was to concentrate on manufacturing, the study gathered information from many other industries also. This data has been re-examined to consider all the industries that took part so that information can be generalised across sectors (although throughout this thesis the study will still be referred to as the MOVE programme). This information has also been published in a book entitled, "*Virtual Reality for Industrial Applications: opportunities and limitations*" (Wilson et al, 1996). Only parts of the MOVE Programme which are the responsibility of this author, or re-analysed by her, are considered in this thesis. The study was to provide answers to the following questions:

- How does industry perceive VR?
- What industrial applications are envisaged for VR?
- How readily will the technology be adopted?
- What are the needs of industry?

The main methods used to answer these questions were to gather information from actual and potential users (e.g. Rover, British Telecom, Crossrail, Vauxhall, Ford, Rolls Royce and Associates, etc.) by relatively direct means, including mass and targeted surveys, in-depth interviews, visits to sites and other laboratories, workshops and semi-structured usability trials. Furthermore information was provided through the literature, internet and professional contacts at conferences and other laboratories and close contact with suppliers and developers of VR systems (Superscape, Division, Virtuality).

3.2.1 The First National Survey on Industrial Applications of Virtual Reality

The First National Survey on Industrial Applications of Virtual Reality (VR) was based on a questionnaire designed to provide initial answers to the questions outlined above. The questionnaire was divided into five sections consisting of open questions and rating scales. The sections and the information obtained by each section are as follows:

(i) Understanding of VR – respondent's definition of VR, how they have heard about it, familiarity and experience of the various systems;

(ii) Advantages and disadvantages of VR – general information about potential advantages and disadvantages of VR technology for UK industry;

(iii) Needs of individual companies – whether the respondent's company have considered VR, applications of interest, benefits achieved or potential, and difficulties associated with implementation;

(iv) General Opinions of VR – agreement and disagreement of common statements about VR; and

(v) Support required – type of support the respondent's needed in order to consider implementation of VR.

A copy of the questionnaire can be found in Appendix I. The survey was the first of its kind to be carried out in the UK and was launched at the VR User Show, Hammersmith, London in November 1993, sponsored by VR News. This event, unlike previous VR conferences, was specifically aimed at industrialists interested in exploring VR for applications. Responses were obtained from 242 attendees with almost a zero refusal rate and the sample represented twenty-one different industries. For analysis purposes it was necessary to categorise the data into general groups representing broad areas of industry as follows: (1) Computing and software (21%) - VR and general software development; (2) Service industries (31%) - architecture, medicine, government and public services, leisure; (3) Finance, sales and marketing (18%) - communications, finance, marketing, sales and retail; and (4) Manufacturing (25%) - manufacturing/engineering, electronics, textiles, chemicals, aerospace, R&D, transport and construction.

Even though, this provided a good response rate, the respondents were essentially self-selected as they were attending a conference specifically about VR. Therefore in order to obtain a wider perspective, questionnaires were also sent to 2000 randomly-selected manufacturing companies. This yielded a total response of 108 completed questionnaires, representing just over 5% return rate. Fifteen types of manufacturing industry were identified including; manufacturing - general, industrial or domestic; materials processing; engineering - chemical, electrical, mechanical, civil; automotive; and aerospace. However the results for the random manufacturing sample will be considered as one group.

3.2.1.1 Findings of National Survey

The National Survey was designed to provide answers to the questions stated at the beginning of this section – 3.2. To the question of what industry understood by VR, definitions produced a set of core keywords that were used consistently by the respondents. In general there was relatively good understanding of VR (although a vast majority of the sample were at a VR conference and therefore would have gained knowledge from this or before they attended). VR technology was described in terms such as three-dimensions, interactive, reality, simulation and visualisation. However the randomly sampled group were notably different. They described VR generally in one or two words, largely simulation (46%), reality (23%), and visualisation (21%) and not many described it as three-dimensional (13%) or interactive (4%), with 10% responses stating that they did not know or gave no answer. This suggested that in general, the random sample of industrialists did not realise the 'interactive' nature of virtual environments, possibly because they had had less direct experience of the technology.

The majority of the total sample's knowledge about VR had mostly come from what they have seen in the popular media such as TV, newspapers and general magazines, but they had not really considered how the technology may apply to their own industry. Knowledge and levels of interaction with particular VR systems were fairly limited and few respondents had actually heard of desktop VR.

A proportion of the total sample (40%) had considered implementing VR, (although this is not surprising considering the way they were contacted). However 12% of this figure were from the random sample group although even this could be an over-estimation if generalising to the rest of industry as only the companies considering VR are more likely to respond to the survey.

The potential application areas being considered included: communication of ideas and concepts, product design, rapid prototyping, training, plant layout and manufacturing process control. The respondents were also asked to choose one that was of specific interest to their companies. Generally the non-manufacturing groups chose 'communication of ideas and concepts' (71%), in terms of, marketing, sales and visualisation. However the two manufacturing groups both chose 'product design' as their specific area of interest. The majority of the random sample also felt that 'training' and 'plant layouts' were of equal interest.

In terms of the benefits to the company envisaged by the sample, most felt that VR could improve communications between departments (53%), and that there would be time savings (60%) and cost savings (50%). In terms of limitations of implementation in their own company, by far the greatest factor stated was the perceived 'cost' (72%) followed by 'the lack of present staff skills' (41%).

Finally in terms of the needs of industry, in general they stated that they required: (i) more information about the technology and how it might be applied; and (ii) demonstrations of actual applications with evaluations of their cost benefits.

3.2.2 Further assessment of potential Virtual Environment applications

Based on the findings of the National Survey, it was decided to carry out further in-depth studies to examine more closely, potential VE application areas. At the end of the survey, the respondents were asked if they were interested in exploring further into VR, nearly three-quarters (71%) agreed and these formed part of the sample for the follow-up studies.

Several data collection methods were used for these assessments including further surveys, interviews and demonstrations. Two surveys were designed and initially sent out to the follow-up sample. One survey was aimed at companies who had considered implementing VR and the second was sent to those who had not considered implementing VR. The survey consisted of open questions. The information obtained from the first survey – to those who had considered implementing VR – included: why they chose VR, did they accept or reject implementation, what system they were using and what applications were being explored. Thirty-one surveys were sent to respondents of the National Survey, of which seven were returned. The information obtained from the second survey – to those who had not yet considered implementing VR – included: why they had not considered VR, if VR had a future, possible systems and applications that would be explored. Eighty-nine surveys were sent to respondents of the National Survey and twenty-nine responses were gained. During this time other enquiries about the survey were being made by companies through articles published about this work in VR News (Cobb and D'Cruz, 1994b) and general enquiries to VIRART. Therefore a one-sheet survey was also designed and distributed to enquirers. This sheet contained a combination of questions from the previous two surveys. Nineteen responses were obtained from this. Therefore the total number of responses from the follow-up surveys was fifty-five. Through these methods different groups were identified: 'Companies who are currently *using* VR'; 'Companies who have considered VR but *rejected it* at this time'; and 'Companies who *have not* considered VR'. These will now be discussed.

3.2.2.1 Companies who are using VR

Seven companies reported that they currently used VR technology and the information provided by them is shown in Table 3.1 (overleaf). Two of these companies were from the automotive industry, one was a manufacturer of domestic equipment, two represented IT/communications industry and the remaining two did not want their sectors identified. The applications reported were: visualisation, modelling, simulation and customer services. Most of these companies had implemented VR within the last year (1994), although three companies did not answer this question. The reasons for using VR were mainly: to integrate with technology performing existing applications (5/7 companies), as well as to perform new applications which are currently not possible without VR (5/7); some companies even wanted to use VR to replace some technologies (3/7). Most of the companies (5/7) were only using one VR system at an investment value of less than £50,000 and all but one company expect ed to gain cost benefits from using VR for their main application.

3.2.2.2 Companies who have considered implementing VR but rejected it

Eight companies reported that they had considered but rejected implementing VR and their responses are shown in Table 3.2 (overleaf). The applications they had considered ranged from product visualisation to sales and marketing in order to gain competitive advantage or facilitate new activities. Most had rejected the technology because they were still finding out what it could do, but others considered the technology not advanced enough for their needs. However all of the respondents did expect to implement VR in the future possibly within the next three years.

3.2.2.3 Companies who have not considered implementing VR

Forty companies who had not yet considered implementing VR were surveyed in-depth and their responses are shown in Table 3.3 (overleaf). Two companies did not see a future use of VR and both of these rejected the technology because they considered it unsuitable for their industries - Food and a Small Manufacturer of domestic goods. Of the remaining thirty-eight companies, twenty-five had not considered VR because they did not know enough about it and eight felt that they had no suitable applications at this time. Other reasons given were concerns about the perceived cost for undefined benefits and the current lack of established industrial use. The majority of these companies could foresee VR being implemented within the next five years (71%) and others expected to be using VR within three years (42%).

Twenty-one companies selected desktop VR as the most suitable type of system for their company and fifteen companies did not know which system would be the most suitable. Only two companies selected both desktop VR and headset systems. Most of these companies did not know whether cost benefits would be achieved from implementing VR and three companies considered that cost benefits would *not* be achieved. Seven companies considered that cost benefits would be achieved through time savings and reduction of errors in design or plant layout activities.

Most of the respondents from this group did not provide further details on possible applications however information from nine of the companies is shown in Table 3.4 (overleaf). One company described suggested that VR could be used to model a showroom to demonstrate their products to customers. The purpose of this application would be to improve sales and the advantages were seen as providing a mobile showroom which allowed more customers to view the product. Disadvantages at present were seen as the high cost of development, lack of easy portability of the headset systems and the relatively poor quality of the current VEs. The remaining eight companies all described potential use of VR in design-related applications. These can be distinguished as either product design (5/8), that is, using VEs to simulate and test a product or workplace design (3/8) that is, to visualise and 'walkthrough' a process or building layout.

All of these companies reported that they were currently performing these activities already, but the use of VR would offer an effective design aid; obtain operator input at the design stage; maximise ergonomic considerations; allow manipulation of complex designs and exploration of alternative scenarios; and be a good sales and marketing ploy. Other perceived advantages of using VR included: potential time savings (6/9), eliminating the need for costly simulation equipment (2/9) and easier training (2/9). Perceived disadvantages of using VR were mostly the initial cost (6/9) and some concern over the suitability of VR technology for specific applications (3/9).

3.2.2.4 Further interviews with industrial users

Other known industrial users of VR technology (located through the literature and personal contact) were also interviewed. In order to maintain confidentiality only the salient points from the interviews are presented here in summary.

The reasons for the companies interest in VR appear to be either to investigate the technology for a particular application (5/10) or to assess the capabilities of VR generally and its potential impact on the company (5/10). The majority of the companies had been using VR for less than two years (8/10) and the remaining two companies had been using VR for research purposes and for development of teleoperations, for more than five years. Applications generally were at an early stage of development and the most common generic application appeared to be 'visualisation' - either of a product or process. Companies had developed applications specific to their own needs but generally (except for two companies) these were being used as 'proof-of-concept' to the rest of the company in order to pursue more investment. Only two companies said that they were using VR for actual 'working' applications. These were 'teleoperation' and 'visualisation of a new plant layout'.

Positive impressions of VR included comments like; "*it has a 'feelgood' factor as everyone can get involved*"; "*it is an important technology of the future*"; and "*it* *is good to keep up-to-date with technology*". Negative impressions were generally based around the technical limitations of the hardware and software which imposed various trade-offs to be made, especially between resolution and interaction. Also the systems did not seem to integrate well with existing technologies in the company and they required specialist knowledge. Finally they felt that there was a lack of example industrial applications to provide guidance in development and evidence of cost benefits.

All of the companies however expected that in time VR would integrate with other technologies, in particular CAD (Computer Aided Design) systems, and at present the technology was 'immature' compared to other solutions and the added value it could offer had yet to be identified.

Suggested improvements to the technology included reduced costs, easier programming and better response to user requirements. Other comments of interest made were suggestions that VR may not be suitable for all applications, but it may add value to CAD technology by offering the facility to model object behaviour which would then allow the user to experience new products. Also it was felt that benefits of VR would not be understood without concrete examples of industrial applications.

3.2.2.5 Findings of further assessments

A total of sixty-five responses were obtained from the further assessments. An additional group of at least fifty people were shown demonstrations of industrial applications and their general comments were noted down. The responses to the studies represented three different stages of VR interest - companies that were already using VR; companies that had considered VR but rejected it at present; and companies that had not yet considered VR. Of the companies who stated that they were using VR the majority of these companies had only been using VR within the last 2 years (10/12 companies) and were currently investigating or demonstrating potential applications rather than developing working applications.

Applications under consideration included product visualisation, modelling and customer services. All wanted to integrate VR with existing technology and expected to gain cost benefits. The general view among these companies was that VR was an important technology of the future and was good for communication of ideas. However, current VR systems were seen as immature, slow, lacking sufficient realism and did not provide adequate integration with other computer systems, specifically CAD. It was suggested that VR system developers should provide industry with the technology it needs rather than "selling promises".

The companies that had considered but then rejected implementing VR had done so because either it was not suitable for their industry or they were still finding out about it. Most, however, did expect to be using it within the next 3 years for applications they said included 'product visualisation' and 'sales and marketing'.

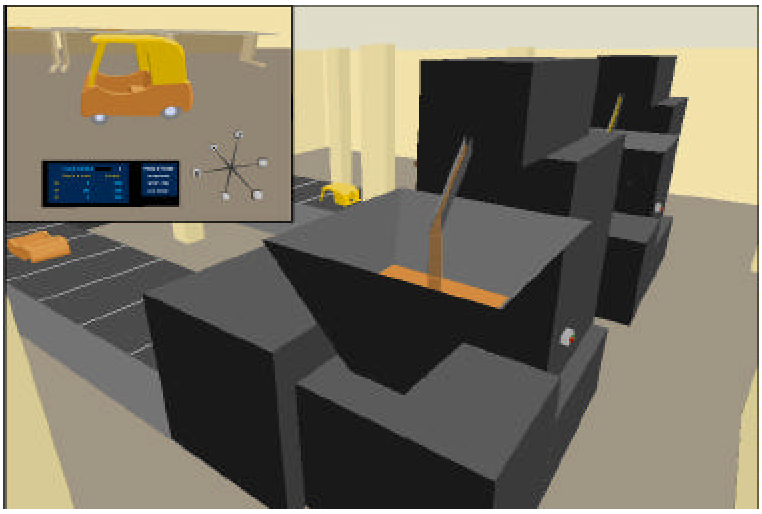
The forty companies who had not yet considered implementing VR either did not know enough about the technology or felt that they had no suitable applications at this time. However only two of these companies did not see a future at all for VR and the remainder expected to be using VR within the next 3-5 years. Where an answer was given (21/40) these respondents selected desktop as the most suitable type of VR system. Not many of these respondents gave details of potential applications but those that did suggested design-related applications (product or workplace) or sales and marketing. Many were uncertain as to whether or not cost benefits would be achieved.

3.2.3 Workshop demonstrations of VR

The conclusion reached through the studies and visits was that VR could not be fully understood without concrete examples of feasible industrial applications. The respondents, even those who were interested in exploring VR technology, had difficulty developing and expressing ideas for specific applications of a technology, of which they had had little experience. In order to address this problem and to provoke more informed judgements of potential, one day industrial workshops were held to provide hands-on experience of different types of VR systems. As part of this a demonstration VE application called the ‘virtual factory’ was built using a desktop VR system to draw attention to specific features of VEs. The workshops had four main aims: (1) to demonstrate the potential of VR systems and the VEs created through these systems; (2) to provide an opportunity for ‘hands-on’ experience for potential users so that they could make ‘informed’ opinions about their own application needs and requirements; (3) to promote discussion between potential users and the developers of VR systems; and (4) to evaluate the views of the user’s about potential implementation of the technology.

Two one day workshops were held and were attended by representatives from a wide range of industries including: engineering, motor manufacture, construction, retail, telecommunications, information technology and computer systems manufacture. The participants spent most of the time working in groups, using the generic VE application developed by VIRART to provide the opportunity to learn about how VEs are created and to demonstrate different key features. Participants were able to take control of operating the VR system and make decisions on how to interact with the VE. They were particularly encouraged to work as a team and to discuss the activities they were performing. The application was called the ‘virtual factory’ (see Figure 3.1 overleaf) and not only showed the potential of each application area but also the added value of integrating a number of applications in one environment. For example, changes made to a product design could be shown to have implications for the manufacturing process, production operations and suitability for the end-user.

Figure 3.1: View of the design room and the manufacturing process of the



'Virtual Factory'

3.2.3.1 The Virtual Factory

The demonstration virtual environment was a manufacturing plant producing toy vehicles for 2 – 6 year old children as can be seen in Figure 3.1. The plastic body and roof components of the product were manufactured at this plant using an injection moulding process. The participant had the facility to modify the product design in size and aesthetic qualities and was able to test the product's suitability for different users (e.g. children of different ages/sizes). The injection moulding process was modelled and could be seen in operation allowing the components to be followed along the production line. The demonstration VE embraced an integrated design-manufacture-test philosophy (for further details see Wilson et al, 1995b) which allowed the examination of a number of attributes of virtual environments applicable to manufacturing.

These included the following:

- modelling in “virtual clay” - dimensioning, reforming and orienting, colouring;
- rapid prototyping through interactive design and test facilities;
- walkthroughs around a factory floor;
- rapid switching of viewpoints, exocentric, egocentric and object-centred
- visual guidance and explanation for operations or maintenance.
- visualisation of several stages in a manufacturing process.
- ergonomics assessment of user-product fit.

There were three main methods of interaction: (1) the function keys on the keyboard which allowed selection of different viewpoints in the VE. Each viewpoint was set at a different place around the factory and some allowed movement control of virtual objects; (2) the mouse which was used to position the arrow cursor on the screen onto objects, allowing interaction at a simple level (e.g. machinery on/off controls); and (3) the spaceball which was used to move the participant around the VE. It responded to different directional control inputs according to the type of object being moved. Also, three different types of movement control were represented in the demonstration: (1) human walking - the viewpoint was set to represent the approximate height of a person and movement control was restricted to two-degrees-of-freedom (forwards/backwards and turning clockwise/anti-clockwise); (2) driving a vehicle - the viewpoint represented an operator seated in a small van or forklift truck and movement control assumed the characteristics of the vehicle; and (3) ‘ghost mode’ - there was complete movement in all six- degrees-of-freedom which could defy all object boundaries.

The virtual factory highlighted three of the main applications of interest to industrialists (as identified through the National Survey and further assessments) - factory walkthrough, visualisation of a manufacturing process and design modification. The demonstration was divided into three stages corresponding with these applications and the participants were given a list of suggested actions to guide them in their experience. For example, the first stage was 'factory walkthrough' and the participant was required to experience movement through the VE using the function keys on the keyboard, to place themselves in different viewpoints and to navigate around the VE using a 'spaceball'. The suggested actions began with "*Look at the factory layout from an aerial view*" then, "*Ride in one of the vehicles*", etc., until all the possible features of 'walkthrough' had been examined. The second stage - visualisation of the manufacturing process - involved the participants observing the injection moulding process in operation using the mouse to activate virtual objects and the third stage - design modification - involved using the spaceball to control a specially designed 'Gyrotool' on the screen to orientate the product design prototype. Then with the mouse and Gyrotool design change facilities were activated and the prototype modified and tested out for different potential users. The consequences of these design changes were immediately reflected in the manufacturing process. At the end of each stage the participants filled in response sheets indicating their impressions of the specific features demonstrated and their utility for industrial applications.

3.2.3.2 Findings of the Workshop

The virtual factory was designed to focus the participant's attention on the features of VR as much as on the applications being demonstrated. They were asked to evaluate the specific examples in the tutorial representing basic facilities of VEs within each mode - 'walkthrough', 'visualisation' and 'design' - and to consider the relevance of these examples to their own potential applications. The results are summarised as follows:

(i) Usefulness of the virtual factory as an industrial tool. The participants considered the walkthrough facility to be useful to industry but were mixed in their opinions of the usefulness of the visualisation and design facilities. Reasons given for these responses were that the ability to walkthrough and interact with virtual models before applying findings to real life could be extremely useful, but that the current user interface is difficult and the system is not suitable for complex modelling. Moreover, although the design facility may be useful for

design/planning activities, it was felt that the demonstration fell far short of current CAD systems in performing these functions. Suggested improvements to make VR facilities more useful to industry were: a better control interface, provision of navigation and visual cues, improved graphics quality and provision of appropriate sound effects. It should be acknowledged that, for all but the first of these requirements, the criticisms reflect the limitations not of VR technology as a whole, but of the virtual factory demonstration itself - deliberately designed to show what a first demonstrator produced by a user company might look like.

(ii) Usability issues. Usability of the VR system and the demonstration VE was assessed in terms of how easy or difficult it was for the participants to perform specific operations. They generally found all design operations including: selection of product components and design features; use of the Gyrotool; viewing the consequences of design changes for production or for the end-user; easy to perform. During factory walkthrough, for manufacturing process visualisation, “knowing where you are in the virtual environment”, and operations to “drive the forklift truck” and “lift the pallet”, caused some difficulty. Using ‘ghost mode’ to fly into the machine and recognise where they were was also found to be difficult.

(iii) Potential industry up-take of VR technology. Considering potential industry up-take of VR technology, participants were given a number of criteria and asked to indicate whether various features would be good enough to encourage engineers or designers (as appropriate) to use VR. They were generally positive about: the use of different viewpoints for all three modes (walkthrough, visualisation and design/test); graphics for the factory walkthrough; manipulation of the virtual objects to visualise consequences of design changes; and use of the Gyrotool to produce rapid design changes. On the other hand, participants were generally negative about the amount and quality of detail for process visualisation and for design purposes. Participants had the opportunity to try a number of different VR systems during a full day, but the majority indicated that they did not consider a head mounted device or even sound effects would improve the use of VR for visualisation or design at the moment; they had more mixed views about

the need for an HMD for the walkthrough facility. Most of the respondents reported that they did feel 'immersed' or 'involved' in the desktop VE and were impressed with its potential, indicating that it would be the most suitable VR system for their industry in the foreseeable future.

(iv) General Opinions. One of the main purposes of the workshops was to provide more informed assessments of VR, which we felt would be achieved through providing some hands-on experience of a VE application. At the end of the day, the participants were asked to complete a questionnaire asking for general opinions of potential up-take of VR within their companies. Seventeen of them had previously completed follow-up surveys and a comparison of their answers revealed that they were generally able to provide more in-depth responses. Most did consider VR to have a future in their company and could foresee specific applications in factory layout, training and marketing. Two companies were currently using VR and six said that they intended to implement VR within the next year. Of the remainder, five companies expected to be using VR within three years and the remaining three companies within the next five years. Foreseen advantages included: cost and time benefits; use as a sales and marketing aid or visualisation tool; and added value to simulation and communication. Thirteen companies expected to gain cost benefits through using VR by reducing development time, making less mistakes in planning and via increased sales. All of these companies considered desktop VR systems to be the most suitable type of system for their company at the present time, although four companies selected headset VR, as well as desktop VR.

A number of suggestions were made about valuable enhancements to current VR systems including: it should be able to handle more detail; it should contain a library of items to make programming quicker/easier; use faster PCs; improve the quality of detail; improve resolution and speed; and better interface with data input/output.

3.2.4 Summary of the MOVE programme

The investigations of the MOVE programme were structured to assess knowledge, views and opinions of a range of potential industrial VR users, from those who had never heard of VR to those companies who were already involved in developing applications. The methods used included: surveys, interviews, visits, demonstrations and workshops and allowed information to be gathered on:

(1) industry's understanding and perception of VR;

(2) applications envisaged for VR;

(3) how readily the technology will be adopted; and

(4) industrial user needs.

Main findings under each of these headings is summarised in Table 3.5.

Table 3.5: Summary of the findings of the MOVE programme

Aims	**Findings**
(1) Industry's perceptions of VR	• technology with potential although current systems are 'immature'
(2) Applications of interest	• still mostly under investigation • product design, visualisation, training, marketing and sales, modelling, communication of ideas
(3) Potential adoption of technology	• within the next 3-5 years
(4) Industry needs	• more information about the technology • demonstrations of relevant applications • reported benefits of VR use • improved technical capabilities • easier user interface • support via user group or network

In general, industrial perceptions and basic understanding of VR were good but they could not consider its applications to their own company's without experiencing some demonstrations of potential VE applications.

The types of applications that were being developed were mostly 'proof-of-concept' demonstrators for internal use, only two companies actually reported 'working' applications. However the types of applications being considered for development were wide-ranging and included: product design, visualisation of manufacturing processes, plant layout, fire protection and occupant behaviour in buildings, testing of interactive processes, training, sales, marketing and data presentation. Time and cost savings were seen as potential benefits which may result from implementing VR, as were competitive advantage, improved service to customers and improved communication. The perceived disadvantages were the initial cost and current immaturity of the technology and the lack of examples of case-studies demonstrating benefits to industry.

As for likely adoption of VR, the companies identified by the study as using VR generally had been doing so for less than 2 years and were either investigating the suitability of VR for certain specific applications or assessing its capabilities and potential impact on the company generally. Cost benefits were anticipated eventually. None of the companies believe that the technology will not be adopted in some form. Of the companies interviewed in-depth who had considered but rejected implementation, the majority were still finding out what the technology could do, but did expect to be using VR within the next 3 years. The companies interviewed in-depth who had not yet considered implementing VR either did not know enough about the technology or felt that they did not have suitable applications at this time. However, the majority of these companies did expect to be using VR within the next 5 years, although most were not certain whether cost benefits would be achieved from its implementation.

In terms of the needs of industry, these included: more information about VR technology and how it might be applied; demonstrations of applications in VEs and of benefits to industry; improved integration with other technology such as CAD and simulation packages; and demonstrated added value to current operations. Suggested improvements to the technology, not surprisingly, included: greater detail of display while maintaining speed for good interactivity; easier programming and file transfer as input and output; improved texture mapping; and reduced costs.

In light of these findings a number of critical questions for the application of VR and VEs, specifically for manufacturing industry were indicated and answers suggested. These can be found in Wilson et al, 1996. However, in particular, four significant barriers to implementation were identified, which are still very much barriers that exist currently.

3.2.4.1 Barriers to implementation of VR

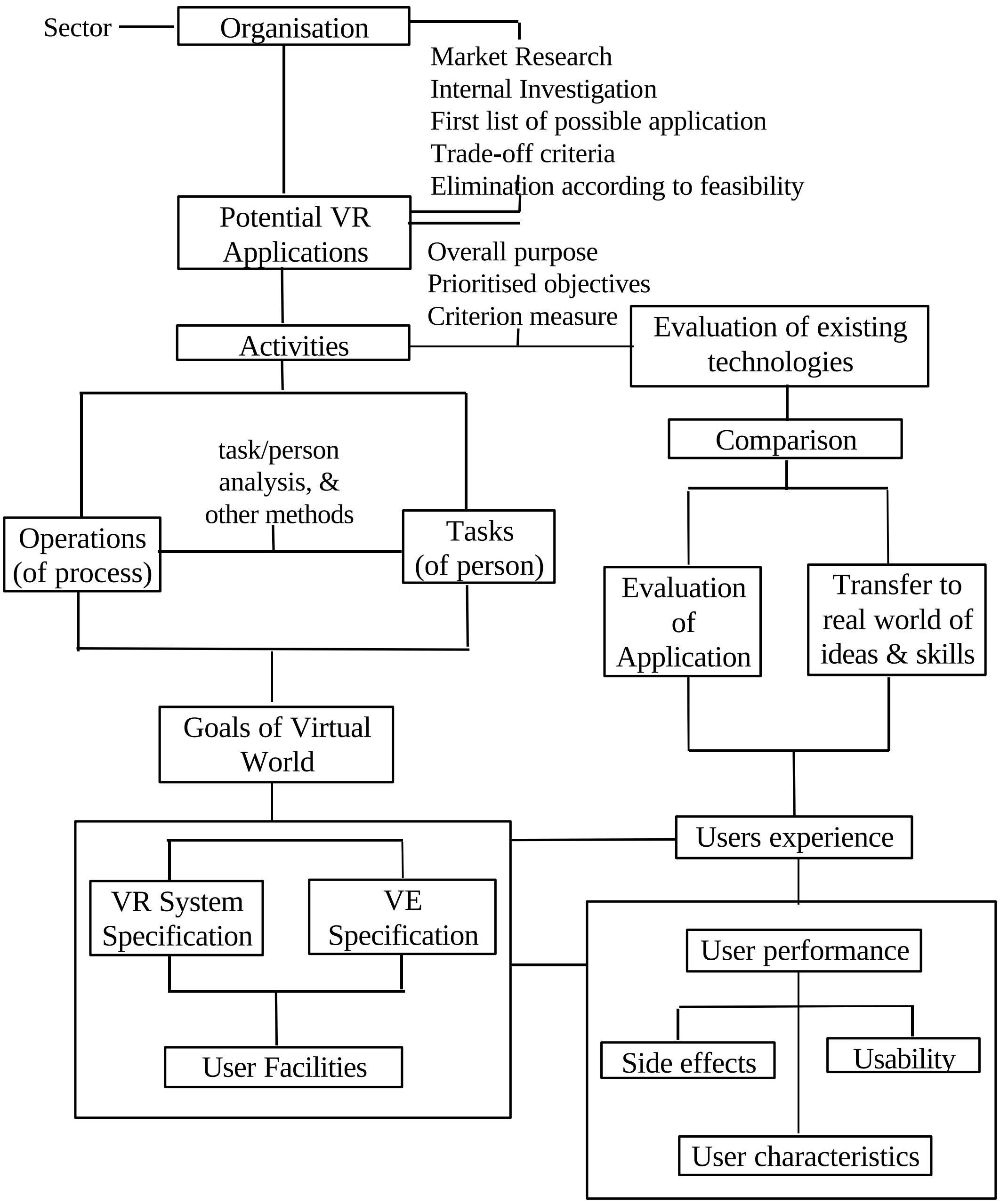
There appear to be four barriers to implementation of VR. The first two of these barriers are the technological limitations and usability problems, (these have been discussed in chapter 2 of this thesis in view of the current status of VR systems). These barriers are continually being decreased as current systems are evolving.

The third and fourth barriers are the lack of examples of working demonstrations of virtual environments and therefore the poor current evidence of added-value from application. These needs - for demonstrations of applications and of their value - are both related to one particular requirement - "*VR use in industry requires a structured methodology to guide application identification, VE building and evaluation against operational criteria*" Wilson et al (1996).

3.2.4.2 Framework for the industrial application of VR

VIRART's experience with a wide range of manufacturing and other industries shows that few of them have a clear and accurate picture of the current nature of VR technology. There is a need for some structured methodology to enable them to identify potential VE applications, produce specifications and turn them into requirements for suitable VEs and then evaluate these applications for added value. An initial attempt at a framework in which to suggest a method is shown in Figure 3.2 (overleaf). This framework considers the process which current VE developers have used to produce their applications. As many of these applications have generally stopped at the ‘build’ stage, VIRART has further extended this process to include the issues involved in evaluation.

Figure 3.2 An initial framework in which to specify, develop and evaluate VE applications (Wilson et al, 1996)



The framework suggests that a good starting point in the development process is the sector in which the company is in, as potential VE applications have more relevance in some industries than in others, e.g. manufacturing companies may be interested in the 'design' and 'test' facilities of VEs, architects would be interested in the 'walkthrough' facilities, etc. Identifying potential VE applications will involve finding out more about the technology in terms of: its reported and demonstrated capabilities; likely changes in the VR market; use by competitors and comparable industries; capital costs; and readiness and willingness of existing personnel to use any systems purchased. From this, a first list of possible applications can be compiled and trade-off criteria identified, based on the perceived costs and benefits of using VR for the given application over present technologies. Costs could be, for example, the greater need for expert personnel, size of system investment or perceived value of existing activities and opportunity cost. Benefits could be the new activities permitted, increased personnel safety, more efficient design processes or better communications. Even if the cost-benefits are expressed in crude terms, this will provide the list of the potential applications to be explored.

Having identified an application the company must consider their objectives for this, which will have consequences for the goals of the VE and for how evaluation will take place, including how they will assess VR as a solution against competing technical solutions. Firstly what must be decided is the overall purpose of the application - for instance in marketing, design, engineering or planning. Then the objectives can be prioritised in order to guide utilisation and division of VR resources e.g. in a marketing application, the graphics receive a high priority but for engineering object behaviour receives a high priority. These objectives, when defined, will also determine the constraints and likely outcomes to be assessed. If possible, each objective should be given a quantifiable rating to aid evaluation of the technology to be used.

The application, which may be relatively complex, should be divided into a number of simpler activities to be clearly communicated to the VE developer. This will involve the use of many techniques (which are considered in chapter five of this thesis). Using these techniques the activities can be further defined into tasks (the step-by-step process of completing the activity) and into operations (the processes required to be modelled and if the consequences of an action need to be shown or are to be part of the activity). Therefore the VE builder is provided with: the application to be modelled; the overall purpose and other prioritised objectives; the activities required to fulfil the application; and the step-by-step analysis of tasks and operations involved. With these identified, the goals of the VE can be established in terms of what must be achieved for the completion of the activities, in view of the possibilities provided by the technology. It is here that careful planning with a thorough understanding of all the attributes, trade-offs and pitfalls of VR is crucial.

The VE developer can then consider the tasks and operations and determine what can be achieved using the particular VR system, the requirements of the design of the VE and the user facilities to be made available. The VE requirements will, in part, determine the choice of VR system and specification of its capabilities and the VE specification will be drawn up in the light of the known technical capacity and facilities of the VR software. Features to be considered include: spatial arrangement - layout of world, location, orientation of objects; VE appearance - lighting and shade, photo-realism level; inter-relationships of objects - boundaries, conflicts, links, priorities; physical properties of objects - mass, friction, restitution coefficient; visual detail of objects - shape, size, colour, texture, visibility, 'solidity'; VE management of detail; and behaviour of objects.

The design of the VE, VR system and facilities allowed will have effects on the user's performance, which together with the user's characteristics, will determine usability and possible side effects experienced. These issues determine the total user experience of the VE, that is, the experience of the user will be effected by five critical variables: content of the VE; capabilities of VE; user interface and usability; user performance in/with the VE; and user expectations, attitudes and affective responses.

When the environment has been developed a process of evaluation must take place. There are many standard evaluation techniques already employed by companies implementing new technologies, including:- examination and manipulation of diagrams, investment analysis, cost-benefit analysis, network analysis, cash flow analysis and strategy analysis. However due to the under-developed nature of VR technology and current technological constraints, actual use is rare and evaluation is still at the experimental stage. However, having defined the prioritised objectives and performance measures before the environment is produced, these can be used to make an initial assessment. In summary, the intended use of this initial framework is to provide a starting point by considering the factors involved in the developing process of VEs. (Revisions of this framework can be found in D'Cruz et al, 1996a)

3.3 Current Application and Research Areas

While the barriers to implementation of VR and VEs identified at the end of the MOVE programme are still very much valid in 1998, the application areas and the development of VE applications has (as is expected) moved on considerably. There are very few people in industry who are not aware of VR technology and the range of applications being explored is wide and varied throughout the world. A list of many of the research and development in academic institutes as well as in industry can be found in Appendix I. In particular, the following areas have received particular attention: military applications; games and entertainment; medicine and healthcare; design, manufacturing and marketing; and education and training (NRC 1995; HITLab, 1997; Hand, 1997).

3.3.1 Military applications

The driving force behind the technology has come mostly from military applications, in particular, for defence and training using high-end VR systems and networking. (Johnston, RS. 1987; Moshell et al, 1990; Alluisi, EA. 1991; Levison and Pew, 1993; Magee, 1993; Johnson and Wightman, 1995; Zeltzer, 1995; Zeltzer et al, 1995; Zeltzer and Pioch 1996; McLin and Chung, 1996) Many of these applications will be discussed in greater detail in the next chapter (chapter four).

3.3.2 Games and Entertainment

The area of games and entertainment has been a major economic driving force for the development of VEs (NRC, 1995). The attraction of VEs for the games and entertainment industry is in making the games more engaging through the three-dimensional graphics and sound and interactions. The actual VEs are very limited in their capabilities (e.g. usually the user has little autonomy and can only move in straight dimensions) however the market has invested much into the peripherals of the systems, mostly input devices, with a result of making the technologies more affordable (Burton, 1993). Therefore compared to all the other application areas, the systems used for VEs are the most widely implemented. Entertainment has made a real breakthrough for VEs and up until very recently the UK has led the field. Generally, in all other fields, the applications are in the stages of research and development, whereas in the entertainment field the technology is commercially exploited and Japan now leads the market (Hand, 1997).

3.3.3 Medicine and Healthcare

Another sector that is focusing on the use of VEs is the medical and healthcare field. Interest is shown by the conferences that are held specifically for this subject including the sixth Medicine Virtual Exposition: Medicine Meets Virtual Reality in January 28-31, 1998 at San Diego, California; the First Joint Conference of CVRMed II and MRCAS III. Computer Vision, Virtual Reality and Robotics in Medicine (CVRMed) and Medical Robotics and Computer Assisted Surgery

(MRCAS) in Grenoble, France (March 20-22, 1997); and the European Conference on Disability, Virtual Reality and Associated Technologies. Also there are a number of commercially available products e.g. the Virtual Anaesthesiology Training Simulator System by CAE Inc., MISTvr - Minimally Invasive Surgical Training and Evaluation System by Virtual Presence and a lot of others, details of which can be found on the internet (HITLab, 1997).

In particular, applications of interest involve improving communication networks, in order to improve diagnosis, planning treatment, provision of information to the patient, provision of treatment and training medical personnel (NRC, 1995). Generally VEs can offer a unique way of information visualisation and manipulation which can be employed to aid surgeons in preparation for operations or in general teaching. The information is generated from MRI (Magnetic Resonance Imaging), CAT (Computerised Axial Tomography) scans, EEG (electroencephalogram), ultrasound and X-rays (Nilan, 1993) and it should allow a three-dimensional representation of the patient's body to be reproduced, in order to understand certain problems and to practice procedures. Applications however are still in the research stage.

Another area within this field is 'rehabilitation'. Work is being carried out to enable people who suffer from cerebral palsy to perform tasks that they cannot usually do (Greenleaf, 1994) and to treat walking disorders associated with Parkinson's disease (Weghorst et al, 1994).

3.3.4 Design, Manufacturing and Marketing

There is much interest in the area of design, manufacturing and marketing (as shown in the findings of the MOVE programme, section 3.2) in terms of product design and visualisation. Throughout the entire manufacturing process, VEs can offer potential benefits (NRC, 1995). For example at the design specification stage, a VE could be used to allow the customer to design exactly what is required. At the detailed design stage, the designers and engineers can communicate more easily their ideas (and show possible show the implications on the manufacturing process, see section 3.2.3.1). The design could also be tested through 'virtual' means and also placed into context, e.g. an engine part could be placed in the whole engine and tested for 'fit' and 'reach-ability'. The design could also be used to plan and schedule the manufacturing process and to market it to potential buyers before it is even made.

One of the best known examples of a VE design application is the Matsushita corporation's Virtual Space Decision Support System (VSDSS) (Nomura, 1992). The system was developed to allow customers to participate in the design of their kitchens. Other reported users include the Rover group and Vauxhall Motors (Wilson et al, 1996) who were both using VEs to design new layouts of their factory plants, and Daimler-Benz and McDonnell Douglas (Wilson et al, 1996) both using VEs to design new parts for engines.

3.3.5 Education and Training

Education and Training are two subjects that cut across a wide number of sectors. They have many similarities and differences based around their scope. Education tends to teach general subjects and training is usually focused on a specific subject. Education in VEs has been focused in two areas - special needs education and mainstream education. Extensive work has been carried out by VIRART in collaboration with the Shepherds School in Aspley, Nottingham, the largest of its type, providing for the needs of children and adults with a wide variety of 'special needs' (Brown, 1996; Brown et al, 1997). This work has involved the teaching of Makaton symbols and life skills, such as crossing the road and shopping. The work has also involved structured evaluations (Neale, 1997). Mainstream education has focused on mainly teaching of science principles, like the Virtual

Physics Lab developed by Bowen Loftin from NASA/Johnson and the University of Houston (Yam, 1993)

In terms of training applications, research in this area is fairly limited even though it was identified through the MOVE programme (Wilson et al, 1996) and by the National Research Council (NRC, 1995) as likely to be an 'early use. of VEs. The existing work in this area is outlined in the next chapter (chapter four). A comprehensive guide to resources on education and training can be found in Pantelidis (1995), Psotka (1995) and Emerson and Revere (1997).

3.4 Discussion

This chapter has presented a view on industry's understanding and perception of VR and VEs and the application areas that are receiving the most focus. Over the duration of this research, views on VR and VEs have changed in the way that, at the start of the MOVE programme (1994) there were still people in industry that were unaware of VR, but by 1996, there were very few people who had not at least heard of the term and now in 1998, more and more work is being reported of research and development in a variety of applications around the world.

In view of the information gathered in chapter two and chapter three it is now possible to state the focus of this research and the reasons behind them. The discussion at the end of chapter two (section 2.6) highlights the inadequacies of many of the components of VR systems. However the components which are associated with desktop VR i.e. a standard computer, monitor, keyboard, mouse, speakers, etc.; have been developed for many years now, so people generally are more aware of how to use them relative to the other systems. Also, this type of system has little of the usability and side-effects associated with most of the other systems, therefore making it easier to concentrate on evaluating the features of VEs. This is important, as it is the VE that needs to be explored, as the nature of VEs are not likely to change, however the technology is (and constantly does so).

Therefore, any research specifically examining the current technology is likely to be out-of-date very quickly. So by using desktop VR, it is possible to reduce the possible negative effects of the chosen system on the performance of the user. Furthermore, the preferred system choice throughout all the industries that participated in the MOVE programme was desktop VR.

At the end of the MOVE programme, the results allowed us to identify a need for some structured guidance for developing and evaluating VEs. The companies requested more information about the technology and in particular, examples of 'working applications' and thus evidence of added value. They required example applications to provide them with some idea of how to develop their own VE applications without wasting resources through an adhoc approach. Therefore, when re-assessing the information from the MOVE programme, an initial framework considering the main features involved in the development process of VE applications, was suggested. This framework was a starting point to develop the framework specifically for developing VE training applications suggested in chapter six.

Finally, 'training' is consistently reported as an application that may find benefits from use of VEs. There is however, limited VE research in this area but a huge amount of information about training theories and methods available to provide an insight into how best to develop and evaluate VE training. Also training is of interest to many different sectors, therefore the contribution of this work will be relevant to many industries. Therefore the focus will be on VE training (VET) applications and the next chapter - chapter four - considers the existing work in the area.

CHAPTER 4: Review of Existing Studies on Virtual Environment Training Applications

4.1 Introduction

This chapter gives an overview of some of the main research and development which already exists in the area of virtual environment training (VET). It is by no means exhaustive but rather considers some of the better known examples which are often cited by other researchers in the area of VET. Few reviews exist in the area (Caird, 1994 and Psotka, 1995) and so this review makes a substantial contribution as a contemporary overview, bringing together examples from industry and academia. The aim is to identify the types of VET applications being explored, the evaluation that has been carried out and the subsequent findings. The first two sections - 4.2 and 4.3 - present case-studies in the military, aerospace and industrial sectors. This is followed by section 4.4, which summarises studies, which have already been carried out to evaluate VEs specifically for training applications. The discussion in section 4.5 evaluates contributions from this literature to identifying the research questions for this study.

4.2 Case-Studies from the Military and Aerospace

For many years the U.S. military, and aerospace in particular, have been exploring the use of VEs for a number of applications, but much of this work has remained unavailable to the general public until recently. Many of the reports have been released and the information is widely distributed through the world wide web. A summary of some of the better known case-studies military and aerospace VET is shown in Table 4.1 (overleaf).

4.2.1 SIMNET (SIMulator NETwork program) and DIS (Distributed Interactive Simulation)

The SIMulator NETwork (SIMNET) project was the first large-scale network simulation using VEs (for details of this work refer to Johnston, 1987; Moshell et al, 1990; Alluisi, 1990 and Alluisi, 1991).

It began in 1983, funded by the Defence Advanced Research Projects Agency (DARPA) in the United States and is still on-going. The project involved networking a number of tank simulators and aircraft together to allow tactical training in realistic battlefields. Presently, there are over two hundred and fifty tank simulators located in Germany, Washington, Fort Knox, Kentucky and a few other places linked together. The trainees control the tank or aircraft simulator while being able to see a 'real' battlefield through a VE. This environment changes in real-time and the battle can be recorded for later briefings. It can also be used by the trainee to focus in on different parts of the battle and to change perspective from one tank to another or to an aircraft.

In 1988, the 'Defence Science Task Force' on 'Computer Applications to Training and Wargaming' recommended the integration of standard training operations with simulations and war games, in order to enhance training.

This resulted in an expansion of the tasks usually trained by tank simulators (i.e. tank hasty attack, deliberate attack, hasty defence and passage of lines) to include manoeuvres, combat support and logistics. These were run over local area networks (LAN) and over long-haul networks (LHN). However, SIMNET had disadvantages because it could only support a maximum of three hundred users, so Distributed Interactive Simulation (DIS) was developed as the next generation of SIMNET to allow for 10,000 to 300,000 users. DIS has been made a standard for communications by the U.S. Department of Defence (National Research Council, 1995) and the IEEE (Isdale, 1993).

The Department of Defence have already used DIS to develop a detailed, 'true' reconstruction of the '73 Eastings' battle which occurred during the Persian Gulf war (NRC, 1995). It is a fully interactive simulation based on actual events and it allows the trainee to explore 'what-if' training scenarios.

SIMNET has shown that large-scale networking of training is possible but what has actually been achieved has been more difficult to quantify. This is because a huge mountain of data has been collected and it is still uncertain what to extract and how to measure group and individual performances (Caird, 1994). Therefore, evaluation is still on-going.

4.2.2 Use of VE training technology for Individual Combat Simulation

The U.S. Naval Training Systems Centre (NTSC) and the U.S. Army Research Institute (ARI) for the Behavioural and Social Sciences, in collaboration with Bolt, Beranek and Newman (BBN) Systems and Technologies, the Massachusetts Institute of Technology (MIT) and Brandeis University, funded a project in June 1991, to review the state-of-the-art of VET technology with regard to developing a simulator for training 'dismounted infantry' (DI) (Levison and Pew, 1993). This simulator would be called the Individual Combat Simulator (ICS) and would be expected to provide three broad training functions, as outlined by Levison and Pew, 1993:

'(1) combat proficiency training (CPT) to train combat units in the execution of various tactical missions;

(2) mission planning and rehearsal (MPR) to develop and refine battle plans and to provide initial training in execution of these plans; and

(3) mission-specific training (MST) to provide practice in executing specific missions.' (P.77)

The project involved identifying ten types of VR technology and then outlining their potential availability on three levels - near, intermediate (3 - 5 years) and far (5 years or more) terms. The results of this are shown in Table 4.2 (overleaf). From existing training documentation, twenty-five tasks and functions were identified to be provided by the ICS and subjective estimates were made of the minimum level of VR technology required to support each of these tasks and functions. Generally they found that even though some technology was available in the near term (level 1 in Table 4.2, overleaf), it was not adequate for training all the required tasks and functions. However it appeared to meet many of the procedural and tactical requirements of combat proficiency training and mission planning and rehearsal and even some parts of mission specific training. Some VET is possible now, but needs advancing to level 3 (shown in Table 4.2) to be fully effective. Other recommended areas for further study included: methods for measuring transfer of training; evaluation of training effectiveness of 'level 1' technology; evaluation of the ability of VEs to provide enhanced or artificial cueing and the causes and effects of 'simulator sickness'.

Table 4.2. Levels of Virtual Environment Training Technology (Levison and Pew, 1993; pp. 21 - 22)

	LEVEL 1	**LEVEL 2**	**LEVEL 3**
(1) Visual Display	Multi-screen	Helmet Mounted	Helmet Mounted

		Display - low resolution	Display - high resolution
(2) Visual Sensing	(none)	Limb and body position	Level 2 + eye position
(3) Auditory Display	Battlefield sounds provided by speakers	Battlefield sounds provided by speakers and headphones	
(4) Auditory Sensing	(none)	Limited speech recognition	Advanced speech recognition
(5) Haptic Display	(none)	Programmable specialised control devices	Programmable general-purpose control devices
(6) Haptic sensing	Joystick, standard control, panel devices	Programmable specialised control devices	Programmable general-purpose devices
(7) Whole-body movement	(none)	Simulated large-volume movement through movement in place	Sensory stimulation involving no motion
(8) Dismounted Infantry Models: Biomechanical Articulation	Perspective view of icon appropriate to stance	Level 1 + low-fidelity articulation of head and limbs	Level 1 + fully animated icon
(9) Dismounted Infantry Models: Influence of Physical Condition	Indication of killed, wounded or operational	Level 1 + movement speed degraded as appropriate	Level 2 + appearance modified as appropriate
(10) Physical Condition of Trainee	(none)	Movement speed degraded as appropriate	Level 2 + artificially induced stress

With regard to transfer of training, Levison and Pew (1993) suggested that studies be conducted to determine the appropriate methods for comparing performance in the VET application with performance in the field (termed as 'performance fidelity') and to assess the transfer of skills (termed as 'training fidelity'). Difficulties lie in the need for a large number of trainees, the time required for training and the definition of performance measures and criteria.

With regard to evaluating the effectiveness of 'level 1' technology, a study comparing VET with current methods is required to determine which approach is the most appropriate, in terms of transfer of training and cost-effectiveness. Further studies would involve comparing 'level 1' and 'level 2' technologies in order to identify the future technology requirements of the ICS.

Also highlighted, in their study of VET, was the potential of VEs to provide enhanced or artificial perceptual cues. These, they considered in four categories: *(1) indication of pedagogical intent*, e.g. use of arrows, flashing objects, bright colours to focus attention to the task; *(2) compensation for physical limitations of the input devices*, e.g. some form of visual prompt to indicate that something is hot or cold to the touch in the real world; *(3) intentional improvement of the cueing environment to speed up the learning process*, e.g. making an important part of an object slightly exaggerated to focus attention; and *(4) intentional degradation of the cueing environment to induce trainees to develop higher level of skill than they would normally*, e.g. changing the weather of a situation to fog, sleet, snow, rain etc., to see how the trainees cope with less visual cues. However, these areas have to be carefully considered in order to ensure that 'performance fidelity' and 'training fidelity' are not being compromised to the detriment of learning.

Finally, simulator sickness, which they define as a feeling of discomfort, including *"nausea and disorientation that occurs while the simulated tasks are being* *performed, plus adverse symptoms that persist (or become initially apparent) after the person has left the simulator"* (Levison and Pew, 1993; p. 70), must also be carefully considered, especially as it has been suggested that greater fidelity and realism in VEs, increases the potential for simulator-type sickness (Kennedy, 1991). As the ICS moves from 'level 1' technology up to 'level 3' technology, this may become a real problem, so studies are required to examine the causes and effects of simulator-sickness in VEs. (Since this work, such studies have been completed, see Wilson et al, 1996; Wilson, 1997; Nichols et al, 1997 and Cobb et al, 1998.)

4.2.3 Officer of the Deck (OOD) Simulator

Leading on from the ICS project discussed above, the U.S. Navy sponsored a further project in collaboration with BBN Systems and Technologies, MIT and Brandeis University. The focus was on training the officer of the deck (OOD) on a submarine (see also Zeltzer, 1995; Zeltzer et al, 1995; Zeltzer and Pioch, 1996a; and Zeltzer and Pioch, 1996b). The OOD is responsible for guiding a submarine in and out of the harbour when it has surfaced. The OOD stands on the 'conning' tower of the submarine and gives rudder, heading and speed commands to the helmsman ensuring that the submarine remains within the channel of deeper water, which is marked by buoys and range markers.

As part of developing the application, the project also involved designing methods for validating and verifying the simulation. These methods were based on Zeltzer's (1992) taxonomy for describing the VE (as discussed in chapter two). This taxonomy consists of three axes labelled 'autonomy', 'interaction' and 'presence' and in terms of this project, Zeltzer and Pioch (1996a, 1996b) defined each as follows. Autonomy, refers to "*the degree to which computational models and processes in the VE behave independently and react to operator input and other changes*". Interaction, refers to "*the logical interface, that is the human/machine interface software of the VE, which are the means by which the user can have an* *effect on objects or conditions in the VE and use the sensory feedback to determine the next course of action*". Presence refers to "*the physical interface determined by the number and quality of the sensors and displays used in the VE*." The requirements of the VET application were then defined using the three areas. For example, in the area of 'autonomy' the VE developer must consider the tasks which should be modelled, the computational models required to resemble these tasks, then the level of detail of these models. For 'interaction,' the VE developer must consider what actions have to be modelled in the VE for the user to effectively learn the task, and, secondly, what input and output devices are required to aid this learning. Finally, in terms of 'presence', which input and output devices are required to promote feelings of being part of the environment. These requirements were determined through close collaboration between the MIT researchers, BBN colleagues and the U.S. Navy officers familiar with the OOD task. The three areas also provided the criteria for validation and verification; in terms of autonomy, the application was evaluated to ensure sufficient amount of detail of models and processes; in terms of interaction, the application was evaluated for providing sufficient understanding of the actions required in the real situation; and in terms of presence, the application was evaluated for system features and VE features which may detract from learning the tasks.

The application was initially validated by domain experts - submarine officers. Their comments resulted in some modifications being made but generally they were very enthusiastic about the potential usefulness for training and mission rehearsal. A pilot study was then carried out with six novices who had little or no navigational experience. After some background instruction and system familiarisation, each performed ten trials of the VET application. The results showed that their performance times and the quality of their performance in the training (i.e. the ability to stay on course), greatly improved with each trial, thus providing further evidence of validity.

Further experiments are planned at MIT to evaluate, in particular, the effectiveness of instructional cues unique to VEs. The U.S. Navy have already initiated the next project which involves developing the prototype OOD simulator for use in the real field.

4.2.4 Officer of the Watch (OOW): the MARS VR Simulator

In a similar project to the OOD simulator discussed above, the Canadian Chief of Maritime Doctrine and Operations (CMDO) requested the investigation of technologies which could produce a simulator to train junior MARS officers the 'conning skills' required by an 'Officer Of the Watch' (OOW). These skills include decision-making and spatial skills needed to obtain the position of other ships in formation and to manoeuvre their own ship appropriately. The task relies on visual and verbal information which is possible with current VR systems. Therefore a prototype using VR technology was developed, called the 'MARS VR Simulator', to be explored for technical issues and to demonstrate proof-of-concept (Magee, 1993).

The MARS VR simulator consisted of an immersive VR system with speech recognition and a voice production system. An iterative design approach was used involving domain experts from Venture, the Naval Officer Training Centre (NOTC) in Victoria, British Columbia, and, also, in one of the three design reviews, appraisal and testing was carried out by four novice trainees. A more extensive evaluation and field trial was also carried out comparing officers that used the MARS simulator with those that had not used the system. A positive transfer of training was reported (Magee, 1993) but no other information was available. However, the conclusions were that the Canadian Navy are now convinced of the benefits of VET for officers of the watch.

4.2.5 The Virtual Maintenance Training (VMAT) Simulator for the U.S. National Guard

The Virtual Maintenance Trainer (VMAT) was developed in 1994, at the Research Triangle Institute (RTI), in North Carolina, to train the National Guard in maintenance and trouble-shooting tasks for the 'M1A1 Abrams' tank, the 'M2A2 Bradley fighting' vehicle and the 'TOW II missile' system (McLin and Chung, 1996). These tasks require the trainee to be familiar with the appearance and location of a number of parts including cables, connectors and line replaceable units. The trainee needs to know how to operate the specialised test equipment used to perform trouble-shooting procedures, as well as to be able to manipulate various switches, knobs and electro-mechanical devices during the procedure. Usually, the trainees are not allowed free access to the real vehicles or even the test equipment, so the VMAT was developed to minimise the time required by the trainee to be spent on the actual equipment.

A desktop VR system with shutter glasses was used and informal anecdotal responses from the field highlighted a very positive response - even the instructors said that they observed benefits in training effectiveness. However no formal evaluation has been conducted, but is planned for the future.

4.2.6 VEs for terrain familiarisation - the Hanchey Army Heliport (HAH)

In 1994, the U.S. Army Research Institute (ARI) for the Behavioural and Social Sciences and the Rotary-King Aviation Unit at Fort Rucker, Alabama, conducted an experiment to explore the use of VEs for terrain familiarisation (for more details refer to Johnson and Wightman, 1995). The VE was a representation of the Hanchey Army Heliport (HAH) at Fort Rucker and the instructional strategy used was self-guided, discovery learning. That is, the army pilots were free to explore the VE and use their own techniques to become familiar with the physical features and information about the flight pattern. Twelve participants were used, all were soldiers from Fort Rucker but with no experience of the HAH. They were divided into two groups - a control that received no previous training but instead explored a virtual map of Arizona and a VET group. A two-part pre-test was administered to investigate their knowledge of the (HAH). The first part required the participant to write their answers down and the second part was delivered verbally by an instructor and the participant responded verbally to the questions. This test was administered again after training, as well as, a questionnaire that required them to place certain objects on a diagram and a further questionnaire on ‘presence’. Finally, in order to examine transfer of training, the VET group were required to carry out two navigation tasks at the real HAH.

The results showed that there were *no* significant differences between the two groups for Part (1) of the pretest questionnaire (t = 0.70; df = 10; p>0.05; two-tailed) or Part (2) of the pretest questionnaire (t = 0.00; df = 10; p>0.05; two-tailed). Therefore there were no significant differences in knowledge of the HAH between the two groups at the pre-test stage. However at the post-test stage, there was a significant difference between the groups for part (1) of the post-test questionnaire (t = 20.78; df = 10; p <.001; one-tailed) and part (2) of the post-test questionnaire (t = 12.38; df = 10; p <.001; one-tailed). Also the scores of the object placement test were significant (t = 7.55; df = 10; p <.001; one-tailed). Therefore the VET group had gained significantly more knowledge of the task than the control group. Scores of ‘presence’ revealed no significant differences (t = 1.08; df = 10; p <.10; two-tailed; Mann-Whitney U = 12; df = 6,6; p <.10; two-tailed). Also, all the participants of the VET group successfully completed the two navigation tasks with no errors and in times comparable to the experimenter. Therefore, as the VET group before training had had no knowledge or experience of the HAH (as shown by pre-test scores) and the post-test scores of the control group showed that changes were not due to - completing the questionnaires; using any VE (the control group used a VE of Arizona before the task); or even changes over time - it was reasonable to assume that transfer-of-training from the VE had occurred. Given these encouraging results the ARI stated that they would further explore the effectiveness of VET and in particular, against other training methods. (One of these further research activities is discussed in section 4.4.4.)

4.2.7 Hubble Space Telescope Repair Training System

Since 1990, Johnson Space Centre's (JSC) Software Technology Branch (STB) at NASA has been exploring VET by developing a number of simulations of space stations and shuttles. At this time, after launching the Hubble Space Telescope (HST) in April 1990, astronomers became aware that the optical system was flawed. This led to NASA/JSC, the Space Flight Division and the Flight Director Office to focus for three years on preparing and training flight controllers, engineers and technicians to repair and maintain the HST, as well as some parts of the Space Shuttle payload bay. VET was used to provide three-dimensional visualisation of the equipment and to allow the flight crew to rehearse procedures. (For more details of this work refer to Loftin et al, 1994; Kenney and Saito, 1994; Loftin, 1995 and Cater and Huffman, 1995).

In 1993, 105 flight controllers received over 200 hours of training in the VE. After the HST mission was completed, a survey was sent to all the trainees for three purposes: (1) to study the effectiveness of the training for enhancing performance; (2) to evaluate the training potential of VEs; and (3) to assess some human factors issues including simulator sickness and other possible side effects. Thirty-eight completed forms were returned and responses were generally positive. On average, the flight team rated the VET as effective for enhancing their job performance and related to previous studies on simulator sickness (Kennedy, 1992), the trainees reported relatively lower rates. The study, which was the first 'working' application of the technology has provided enough evidence for NASA to believe that VEs are beneficial for supporting training of complex procedural tasks. In particular, they can provide training to trainees who currently receive little or no experiential preparation for their missions.

4.3 Case-studies from industry

There are very few reported examples of VET applications in industry. Reasons probably include: maintaining confidentiality so that competitors are unaware of their activities especially if they are providing cost and time saving solutions, but also an actual lack of applications which have been developed because of the uncertainty of added benefits. Some of the better known applications are listed in Table 4.3 (overleaf). In many of the cases, the author has visited or made personal contact with the companies involved and carried out informal interviews.

4.3.1 Motorola - Training of line workers to operate a Pager Robotics Assembly unit

Motorola has manufacturing plants all around the world and each year the equipment in some of the plants is up-dated or a new factory is built. Motorola could hire new employees with the skills necessary to run the new equipment but they are committed to re-train employees rather than replace them. Since many of the factories use robotics instead of manual labour, Motorola wanted all its 'associates' to understand how a robotics manufacturing line operates. The usual method for this was to send the associates to a three day class - An Introduction to Advanced Manufacturing Concepts (MFG451), at Motorola University in Schaumburg, Chicago.

The course consists of a lecture, classroom activities and hands-on activities on a five station manufacturing line. The training lines cost approximately one million dollars to build and they require maintenance. They cannot be easily transported to other facilities so outside the Motorola University, no hands-on activities are carried out. Instead trainees watch a video of equipment operations.

In July 1994, Art Paton, the instructional design manager at Motorola University, in collaboration with Nina Adams of Adams Consulting began to explore the possibilities of developing a VET application which would provide the trainees with 'hands-on' experience of the Pager Robotics Line. (For more detailed information refer to Adams, 1996 and summarised in CyberEdge Journal, 1994; Wittenberg, 1995; VR News, 1996a; VR News, 1996b and Wilson et al, 1996.)

The aim of the project was to compare the performance of groups trained by a VET application with groups trained in the real environment. The VET application was developed using standard instructional design stages - data gathering, design, development and testing. Information was gathered through reviewing all the current course material, the equipment manuals and video taping the operation. Motorola and Adams in close collaboration agreed which activity from the existing course would be modelled. This activity consisted of the tasks of starting-up, setting-up, running and shutting down the equipment on an automated assembly line without any job aids or checklists. The learning objectives to be measured and the evaluation criteria were then agreed. The application took four months to develop using a full time programmer, a part-time analyst/project manager and a domain expert.

Twenty-one associates from the manufacturing facilities in Northern Illinois were selected for the evaluation trials. They were divided into three groups. All the participants were given the regular classroom instruction at the start of the session. Immediately afterwards, one group went directly to the manufacturing laboratory where they used the existing approach to learn how to start-up, run and shut-down the manufacturing line. This approach involved following a checklist for each task. The second group were taken to another room and began training on a desktop VR system and the third group were trained by the same system but with a head mounted device (HMD) added.

After about an hour of training, each participant was taken into the real manufacturing lab and asked to perform the task without the aid of any form of checklist. Their performance was measured by errors made, through observation by a certified instructor. After the task they were also interviewed.

The results showed that, on average, in the most complex parts of the activity (start-up and set-up), the participants who had learnt in the real laboratory made an average of fourteen errors where both VE groups made an average of one error. Generally, the participants who were in the VE groups appeared to be doing as well as (or even better in some cases) then the ‘control’ group.

The project is regarded as the first industrial study to actually evaluate a VET application against an existing training method and to show some transfer to the real situation (CyberEdge Journal, 1994; VR News, 1996a). While objective results were minimal, the subjective opinions of the instructors themselves has encouraged Motorola to pursue VET further, developing a module to be implemented as part of their existing training programme.

4.3.2 Landis & Staefa - Training of engineers and technicians in maintenance of heating, ventilation and air conditioning systems

Another VET project which involved the Adams Consulting Group was carried out for Landis and Staefa (previously known as Landis and Gyr), a company involved in air conditioning systems (see also VR News, 1996a; VRET, 1997 and http://www.adams-consult.com.html). They wanted to reduce the time it takes to train engineers and technicians to maintain heating, ventilation and air conditioning systems. They felt that VET could provide the trainee with some experience of a number of frequent scenarios.

The VET application begins in a virtual office with a phone ringing. The trainee picks up the phone and hears a problem from a client. They must then use the available resources i.e. various standard 'virtual' tools like ladders, torch etc., to discover the cause of the problem. Pilot testing of this application is due to involve 20-30 trainees and the same number for a control group. However information about this test has not been made available as yet.

4.3.3 AMOCO Truck DriVR - Training of truck drivers in defensive driving in emergency situations

AMOCO in collaboration with Bravo! Multimedia have developed a VET application called 'TruckdriVR' which allows truck drivers to learn how to better handle road emergencies and other hazardous driving situations (for further details see VR News, 1996a; I/S Analyzer, 1997; VRET, 1997 and VRU, 1997). This type of training is very important as the trucks often carry hazardous materials through highly populated areas.

The project took a year and involved an instructional designer from Bravo! Multimedia, two to three designers from AMOCO and two to three programmers from ‘Sense8’. The scenarios were emergencies which had a number of outcomes. The drivers were tested for ‘SITE’ (Search, Interpret, Time and Execute) for each of the scenarios. Initial response by the truck drivers has been very positive and field trials are currently being carried out.

4.3.4 Nortel - Training of attendant console operators in the features of phone systems

Nortel has developed a VET application which can be delivered on CD ROM to their customers - attendant console operators - to allow them to learn the skills required for their M2250 system through self-directed, discovery learning (see VR News, 1996b or NORTEL_VR@NT.COM). The tasks include setting up the console, identifying the major components, activating basic features, activating convenience features, identifying the keys and indicators on the console and responding to calls. The course takes approximately six to eight hours to complete.

Each trainee also has a ‘Passport Diskette’ which records individual progress. Once the course is completed the trainee returns the Passport Diskette to Nortel for recording in a registration database and a ‘Certificate of Completion’ is mailed to them. Internal evaluation has been completed but further information is not available.

4.3.5 Volkswagen Dealer Training System - Training assessment of service engineers in performance of car repair and maintenance tasks

Funded by Volkswagen, Logicom Sound and Vision have developed a ‘VR dealer training system’ (see VRET, 1997 or http://www.vrs.org.uk/VR

/news/current/news0009.html). The aim of the system is to test potential engineer recruits for the necessary qualifications without using a physical mock-up or real car which can be expensive and time consuming. The application took six months to complete and involved two people from Volkswagen, three programmers/graphic artists/modellers from Logicom and two technical people from Superscape. The user explores a generic car and hot spots appear indicating a place which can be closely inspected. Pictures are brought up in progressive detail and various parts of these can be tested, e.g. the brake lights, fog lights, reversing lights etc. using a number of tools to solve problems or change parts. Feedback in the form of text is delivered afterwards indicating the actions taken to solve problems and the subsequent cost. The better employee is the one who finds the cheapest way to fix the car, e.g. the engineer may have replaced all the fuses, which is quite costly, without checking which ones are at fault, which is cheaper. The application has undergone internal evaluation and is now installed in approximately 350 locations.

4.3.6 Southwestern Bell Corporation, SBC TOWN - Training of telephone engineers in fault-finding in telephone lines

Funded by Southwestern Bell Corporation (SBC), EDS developed a complex VET system for training telephone engineers in locating faults in telephone lines, (for further details refer to EDS' websites – http://eagle.xweb.eds.com/industries/communications/industry_profile/feature_stories/cm_virtual_city.shtml). The traditional lecture and laboratory method had a number of disadvantages, e.g. it was too expensive, limited to a few locations, had too slow a throughput of engineers, the trainees had different backgrounds and experience and motivation levels were difficult to keep high. Therefore they were looking for an application that was cost-effective, portable, could minimise the training schedule, allowed the trainees to move at their own pace and made training more interesting.

The resulting VET application was called 'SBC Town' and could be used to teach technical concepts and skills, logic and problem solving, environmental issues and safety practices, as well as cost control. Eight scenarios are given to the trainee in the form of simulated 'work tickets' which describe the customer's problem. The trainee has to navigate through the town to locate the fault by using what they have already learnt from introductory sessions. The trainee also has a virtual tool box which contains all the necessary tools for fixing faults. The application also encourages them to find the most cost effective method of fault finding by making them aware of what resources they have already used and what they have left.

After demonstration of SBC town, Southwestern Bell felt that the engineers who would use the application would receive *"better, more efficient training that will enhance their performance"* and produce training-related savings. Therefore, they are looking at expanding the use of VET to other tasks.

4.3.7 Fluor Daniel, I.C.I and U.K. Health & Safety Executive - Training of operators in safety procedures in large petrochemical plants

Funded by Fluor Daniel, ICI and the Health & Safety Executive (HSE), VR Solutions Limited developed a VET application consisting of part of a petrochemical plant (see VR News, 1996a). The aim was to create awareness of the dangers of accessing vessels for maintenance and cleaning procedures without using the correct restraining and respiratory safety equipment. The trainee enters one of the vessels and if they have not performed the correct procedures before entry e.g. permission to enter, requisition of protective equipment, etc., then a simple visual representation of harmful chemical vapour is released. Movement is slowed down and their 'virtual' vision begins to degrade in order to demonstrate the effects of the vapour and the limited time for evacuation.

4.3.8 Electricity Supply Commission Of South Africa - Training of operators in safety procedures on high voltage systems

Funded by the South African Electricity Commission, Fifth Dimension Technologies (5DT) have developed a VET application to train safety procedures on high voltage systems (see VR News, 1996a). This application, called the 'VR High Voltage (VRHV) Training System' allows the trainees to experience and understand real-life scenarios in this extremely hazardous environment. In particular, the operators can rehearse the safety procedures thus reducing error rates so that the operator is not at risk or the electricity supply is not disrupted in any way.

4.3.9 Rolls Royce and Associates Ltd - Training of submarine personnel for maintenance tasks

Since about 1992, Rolls Royce and Associates Limited (RRA) have been exploring the use of VEs for various applications (Morrissey, 1996; Wilson et al, 1996; VRET 1997). An early prototype model was developed in collaboration with VIRART, to train two levels of submarine personnel - expert and novice - to familiarise themselves with the a compartment layout and to carry out maintenance procedures. At the expert level, instruction is given in the form of procedural statements, e.g. *"The cover cannot be removed until all the bolts are undone"*, as well as information about schedules and product number. At the novice level, the whole maintenance exercise is given in the form of precise checklists, e.g. *1) This is the cover of the motor generator set. 2) It gives access to the fine brush gear, etc.* This work generated enough interest for RRA to pursue further VE projects (in particular in the area of radiation dose management).

4.3.10 NCR (National Cash Register) Corporation – Training of basic replenishment tasks on an automated teller machine (ATM)

In 1994, the NCR (National Cash Register) Corporation approached VIRART at the University of Nottingham to consider the use of VEs within their company. A number of applications were considered and training was chosen as an initial application. The main aim was to develop a VET application that would highlight the key features of VEs. Two replenishment tasks were chosen – changing a printer receipt roll and refilling the money drawers. The target users were chosen as trainee engineers who would use the virtual ATM to practice procedures. Therefore no instructions were included, instead the trainee was required to perform the task as they would be expected to in the real world. (Further information can be found in Eastgate et al, 1997)

A number of domain experts validated the VE and then it was presented to representatives of the company from different departments. Feedback was very encouraging with other departments able to see potential for their own applications. In particular, the marketing and sales department could see the potential for providing a mobile showroom of their products.

4.4 Examples of existing evaluation studies on VET applications

Just as the potential of VET is beginning to be explored by different sectors, so, too, are academic institutions and research and development centres beginning to study various features of VEs and VR systems, specifically for training. However, these studies are few and confounded by many factors, the major one being the constantly changing technology. The better known studies are shown in Table 4.4 (overleaf).

4.4.1 Training of procedural and navigational tasks - Regian et al, 1992

This research was supported by the U.S. Air Force Office of Scientific Research and involved the Armstrong Laboratory on Brooks Air Force Base in Texas and the Galaxy Scientific Corporation at the Lackland Air Force Base in Texas. The purpose of the study was to explore the 'instructional' potential of VEs and to describe the direction for future research. Thirty-one participants from a temporary employment agency were recruited. They ranged from 18 to 35 years of age and were paid for twenty hours of participation. Two VEs were developed for the experiment. A virtual console was used for training a small-scale procedural task and a maze with three levels and four rooms in each level, to train a large-scale navigation task.

For the procedural training, the participants were randomly divided into two groups. Both were trained by the virtual console to perform a complex seventeen-step procedure which was highlighted through visual prompts (green highlights on the buttons and knobs to be activated). One group however was given a 'meaningful task description', that is, as the participant interacted with each button and knob, the instructor would give them information like the name of the button and its function on the console. The other group was given no explanations but just told which button to activate next. Both groups performed the same testing phase. They were given three minutes to perform the same procedure five times with no external prompting and with a two minute rest period between each performance. All the participants were able to complete the five trials within the three minute time period and there were no effects on accuracy or speed of the performance of the participants based on whether they received meaningful instruction or not.

For the navigational training, all the participants were given three guided tours of the maze starting from different initial positions, in the form of verbal commands. Then they were given an hour to freely explore the maze. Testing involved performing three tours (different from the guided tours) and the aim of the participant was to get to another room from their initial position, visiting the minimum amount of rooms. The results showed that generally the participants were able to learn the virtual maze and most of them were able to perform the task against the criteria set of minimum rooms (estimated by statistical means considering random performance distributions).

The experiments provided some evidence that participants can learn 'spatial-procedural' and 'spatial-navigational skills' in VEs, however further experiments are planned to consider whether these skills can be transferred to the real world.

4.4.2 Training of psychomotor skills - Kozak et al, 1993

This work was carried out in the Human Factors Research Laboratory, University of Minnesota, Minneapolis and is noted as the first study to consider the issue of transfer of training from VEs to the real world. Twenty-one participants were recruited from the University environment, ranging from 18 to 59 years of age. They were divided into three groups - a control group, a VET group and a real world group. The task to be performed involved a simple pick and place task, where the participant was required to move five cans from an initial position to a target position and then back again to represent one trial. The control group received no training, the VET group were given twenty minutes to familiarise themselves with the system and then performed thirty trials of the task in the VE, and the real world group performed thirty trials in the real world. Then, all the groups performed thirty trials in the real world while being timed.

The results for training performance showed that applying a two-way repeated-measures analysis of variance (with trial block as a within-subjects factor and training group as a between-subjects factor) showed no significant differences in performance times between the training groups (F (1,12) < 1.0). There was a significant decrease in response times across blocks of five trials for both groups ($F = 14.914$; $df = 5,60$; MSe $= 0.153$; $p = 0.001$) but there was no significance between the interaction of the training group by blocks ($F = 1.583$; $df = 5,60$; MSe $= 0.153$; $p = 0.179$). In terms of performance of the task, again applying a two-way repeated-measures analysis of variance and then a Tukey's HSD showed some significant differences in the first ten trials between between the real world group and the control group ($p = 0.003$) and the real world group and the VET group ($p = 0.025$) but no differences between the control group and the VET group. Therefore, it was concluded that learning did not transfer from the VE to the real world as the VET group were only as good as the control group.

However, there were certain aspects of the VET system observed which may have resulted in this lack of transfer. In particular, the users had a virtual hand in the VE, which was possible to lose if the user went beyond the tracker's reach. They then had to relocate their hand, thus distracting them from the task. Also, there was some time lag between the user making a grasping action in the real world and the virtual hand replicating this. In the real world such actions like grasping are fairly simple and have become an automatic response for most people. However in a VE these actions are complicated given the current state of the technology, as the user has to use generally less then ideal input devices to carry out an action. For example, the user may have to mentally link clicking on the left button of the mouse with the actions of 'pulling' a door handle, 'picking up' a cup, 'pushing' a chair, etc., these links may not be automatic, therefore making these simple actions quite complicated.

4.4.3 Training of psychomotor skills - Kenyon and Afenya, 1995

This experiment was carried out to re-examine the Kozak et al, (1993) study discussed above, by using the same pick-and-place task but a different VR system. The work was funded by a National Science Foundation grant (IRI-9213822) and involved the Department of Electrical Engineering and Computer Science at the University of Illinois in Chicago. Twenty-four participants from the University environment ranging from 24 - 45 years of age were recruited. The pick and place task was adapted to have two levels of difficulty - ordered and random. The five cans were colour coded and had to be moved to their matching colour in the target area. At the first level, the colours in the initial and target areas were aligned in order and at the second level the positions in the target area were randomly assigned. There were also two conditions of the VET application - 'cursor attached' and 'cursor detached'. The cursor refers to a position prompt in the VE which was either 'attached' or 'detached' to the participant's hand. This was to determine any difference between 'proximal' and 'remote' manipulation of objects. The participants were therefore divided into four groups - VET with attached cursor, VET with detached cursor, real world with attached cursor and real world with detached cursor. Training in one environment meant that they were tested in the other environment and all the participants performed both levels of the task.

The results for completion times of the tasks performed in the VE were significantly longer than tasks performed in the real world. The average completion times for the ordered and random disk arrangements for all subjects were 19.28 seconds and 21.21 seconds (respectively) in the VE and 7.37 seconds and 10.01 seconds in the real world. Also there was a significant difference between the VET (attached) group and untrained participants (where untrained data was represented by the groups that performed the real world task initially) but not between the VET (detached) group and the untrained participants. The conclusions were that some transfer of training could be shown but only under certain conditions. The differences in times were explained by the large differences in sensory information between the VE and the real world - in particular the lack of tactile feedback in the VE. With a delay (detached cursor), the performance of the participant was no better than the untrained group. The study reached a similar conclusion to Kozak et al (1993), in that possibly a task which requires the participant to form a strategy rather than to just acquire motor skills would greatly benefit from VEs. Additional experiments are planned to explore various visual features e.g. shadows, textures, accommodation etc. and their impact on performance.

4.4.4 Training in navigation skills - Witmer et al, 1996

This experiment was collaborative work between the U.S. Army Research Institute (ARI) Simulator Systems Research Unit in Orlando, Florida, and the Institute of Simulation and Training (IST) at the University of Central Florida in Florida. The aim of the experiment was to evaluate the transfer of route knowledge from a VE of a complex office building to the actual building. Sixty participants from the University environment were recruited. They all studied route directions and landmark photographs and then were divided into three groups - a VET group, a real world group and a 'symbolic' group. The VET group rehearsed the procedure using the VE model, the real world group used the actual building and the 'symbolic' group verbally rehearsed the directions out aloud and had the landmark pictures to view. Measures of the time to complete each rehearsal trial, number of wrong turns; number of collisions were noted and after three rehearsals, the VET group received the Simulator Sickness Questionnaire (Kennedy et al, 1993) and Presence Questionnaire (Witmer and Singer, 1994). All participants then completed the procedure in the actual building and knowledge of building configuration was measured by getting the participants to estimate the distance and direction of their initial positions to another position.

The results showed that the VET was able to transfer more route knowledge than just 'symbolic' or verbal rehearsal, but less than rehearsal in the actual building. However, type of rehearsal had no effect on configuration knowledge. The conclusions were that VEs can train navigation skills in complex buildings and, therefore, should be considered whenever the building in the real world is unavailable for training.

4.5 Discussion

The case-studies reported in this chapter represent the research and development being carried out in the area of VET. In general the types of VET applications being developed involve combinations of: (1) familiarisation of a product/process/place; (2) rehearsal of procedures for repair and maintenance; and (3) problem-solving/trouble-shooting scenarios. These applications appear to be mostly demonstrations of proof-of-concept and therefore little evaluation has been carried out beyond initial feedback. The feedback though has been generally positive and has initiated further work in the area. Of the few evaluation studies that have been attempted, the main objectives have been to measure the effectiveness of VET in transfer-of-training. However, these evaluations have been confounded by the many uncontrolled variables which could account for learning and performance effects. In particular, problems have occurred because of the large number of configuration options for VR systems. This also makes it difficult to compare results.

The potential benefits of VET for the military and aerospace though are fairly obvious as both sectors deal with training in situations which are remote, hazardous and/or generally inaccessible. For many years, this has meant that some form of simulation has been required to provide trainees with a resemblance of 'hands-on' experience before they carry out their missions. The use of VET is a natural progression from simulators, as in comparison they are cheaper to produce and reproduce, more flexible to design changes, more portable, can provide support for other applications (such as product/process design) and brings the action much more closer to the trainee. However evaluating these benefits is problematic because generally (with the exception of the VMAT) high-end VR systems with different configurations of the modes of display and interaction have been used. This has made it difficult to extract information about the trainee's performance from the performance of the VET application. which makes evaluation of effectiveness and transfer-of-training extremely complicated. However, what has been highlighted are the limitations of the current VR systems to fulfil all their 'simulation-based' training requirements and the real need of methodologies for effective evaluation.

In comparison, the VR systems being used by industry are, generally middle- to low-end systems where VET is an extension of lecture room and laboratory sessions. The VET applications tend to be demonstrations in order to justify further investment, so little evaluation apart from initial feedback has been performed. However the industries which have begun developing applications have all concluded that this initial response has led to further exploration of VET.

Therefore, there is still a real need to evaluate VET for effectiveness and transfer-of-training. Of the few existing experimental studies in VET, their use of high-end VR systems have resulted in usability problems and side effects which have made it difficult to show any real benefits specifically from VEs. Also the researchers have noted that the use of VEs may only be effective for certain tasks in certain conditions.

The success of a VET application appears to depend on the type of task chosen, the configuration of the VR system, the design of the VE and the methods of evaluation. Therefore, the whole development process of a VET application has to be carefully considered in order to develop successful evaluation methods, (this is further discussed in chapter six.)

Therefore, the findings of the case-studies reinforce the reason why this research will focus on desktop VR systems (as discussed at the end of chapter three section 3.4). This is because little experimental work has been performed exploring the use of desktop VR systems for training, whereas the case-studies of industry use of VR and VEs have suggested that in the short-term, desktop VR is likely to be the preferred choice of system, as it requires the minimum amount of investment and has little of the problems associated with other systems (see chapter two). Also the type of training tasks that the existing studies have examined are those that involve either navigation or psychomotor skills. Some value for use of VEs for navigation skills has already been shown (Regian et al, 1992; Witmer et al, 1996) but there has been little evidence to show the benefits for psychomotor skills (Kenyon and Afenya, 1995). Therefore this research will focus on training of psychomotor skills using desktop VR.

Finally with regard to the aims of evaluation of the case-studies, generally the first issue is to examine the effectiveness of the application, with particular reference to the trainees and secondly to explore the issue of transfer-of-training. Therefore these will be the evaluation aims of this work.

The next chapter considers existing theories and methods on training and evaluation for their possible recommendations on evaluating virtual environment training applications.

Chapter 5: Theories on training and evaluation

5.1 Introduction

This chapter presents theories on training and evaluation where they are relevant to this research. The examination of existing research and development of virtual environment training (VET) applications (in chapter four) revealed little information but did highlight the many potential problems with evaluation of VET. These were partly based on the configuration of the VR system used and the design of the VET. In order to develop effective evaluation techniques, it is not only necessary to consider the existing methods of training evaluation, but also to understand the whole training development process. There is a vast amount of literature in the area of training, much of which is beyond the scope of this thesis. Therefore the intention of this chapter is to not examine each area in huge detail but rather to highlight the major influences in the area in order to guide the development of VET. The chapter begins by defining training and evaluation in section 5.2 which leads to a look at the approaches of training design in section 5.3. Then sections 5.4, 5.5 and 5.6 consider the main stages of the development process - training needs analysis, design and development of training and particular attention is aimed at the evaluation stage. Finally, in section 5.7 the impact of computers in training is discussed along with some of the technologies associated with training methods. This is followed by a discussion of the main points.

5.2 Definitions of training and evaluation

Goldstein (1993; p. 3) defines training as: "*the systematic acquisition of skills, rules, concepts or attitudes that result in improved performance in another environment*". Therefore it can be seen that training involves understanding how to design an application to be delivered in a systematic form; understanding different theories on acquisition of skills, rules, concepts and attitudes; and developing methods that can reveal an improvement or change in performance in another setting.

The evaluation of training is defined by Goldstein (1993, p. 181) as "*the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to selection, adoption, value and modification of various instructional activities*". This suggests that evaluation is also structured and it involves gathering appropriate information about training to make accurate decisions on the selection of the various features of the application; the likelihood of implementation; the actual value of the application; and to identify any modifications that may be required.

Therefore these definitions of training and evaluation highlight the need to understand the systematic development of training and a number of psychological theories that support this development process. These approaches are discussed next.

5.3 Approaches to training development

Patrick (1992) suggests that there are two approaches to development of training - the 'psychological approach' and the 'systems approach'.

The 'psychological' approach of training considers the various theories which exist which contribute to the procedures of analysing a job, designing the training and evaluating the outcomes. For example, as Goldstein's (1993) definition of training suggest (section 5.2) training involves the acquisition of skills, rules, concepts or attitudes. Therefore it is necessary to consider the many theories that relate to these concepts. Comprehensive reviews of many of the existing theories and references can be found in Patrick (1992) and Goldstein (1993) and these texts are used frequently throughout this chapter.

Patrick (1992) comments on how the theoretical approaches to learning have changed over the years, in particular three shifts in perspective can be noted. At the beginning of the century until the 1950s, there were two schools of thought based on the 'behaviourist' approach. The first stated that learning occurred through the principle of 'contiguity' or association (citing Guthrie and Tolman) and the second was based on the principles of reinforcement of stimulus-response connections (citing Thorndike, Hull and Skinner). The next period was between the 1960s and 1970s, which was the start of cybernetic theory which stimulated the idea of information processing models of skill (citing Welford, 1968) and also the importance of feedback and control of skilled performance (citing Adams 1978; Annett, 1967; and Bilodeau, 1966; and discussed in detail in Stammers and Patrick, 1975). The final period was influenced by the emergence of cognitive psychology which changed the view of the trainee from being a 'passive' learner to an 'active' learner, in the sense that the trainee has different strategies, perspectives and interpretations on the situation that can promote learning (citing Schank and Abelson, 1977; Bower, Rumelhart and Norman, 1978; Black and Turner, 1979; Rabbitt, 1981; Anderson, 1982; Broadbent, 1987). Some of the main theories that contribute to this area are discussed in context throughout this chapter.

The 'systems' approach to training (SAT) is an attempt to provide an overview of the training process and the 'subsystems' involved. In particular it highlights two important issues: (1) training is a system that interacts with other systems within an overall organisation, e.g. personnel selection and ergonomics - all three may provide a potential solution to a performance problem, in isolation or in combination and therefore need consideration; and (2) the development process of training can be viewed as a system which can be analysed into subsystems of different functions which are performed in the process. This has given rise to the production of Instructional Systems Development (ISD) models which are used as tools for developing training programmes (Patrick, 1992; Goldstein, 1993). There are many different types of ISD models, Logan (1982) stated that Montemerlo and Tennyson (1976) had found more than a hundred manuals containing models since 1951 and Andrews and Goodson (1980) reported over sixty. Reigeluth (1983) provides a review of eight of the better known models. Three examples of such models are shown in Figures 5.1, 5.2 and 5.3 (overleaf).

Figure 5.1: An instructional system (Goldstein, 1993; p.21)

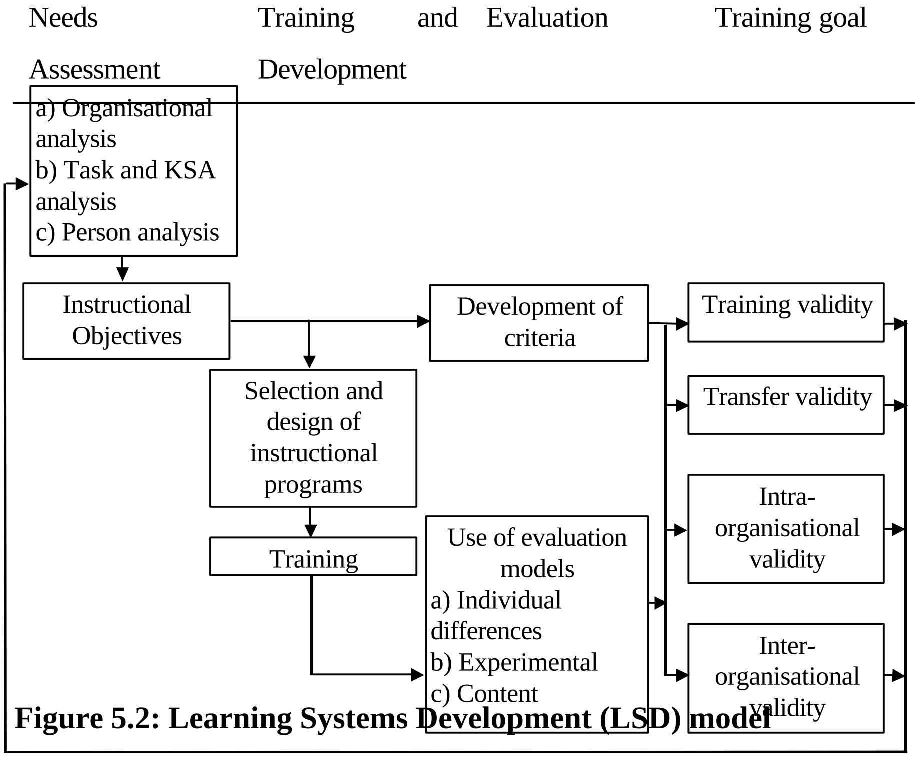


Figure 5.2: Learning Systems Development (LSD) model (Patrick et al, 1986)

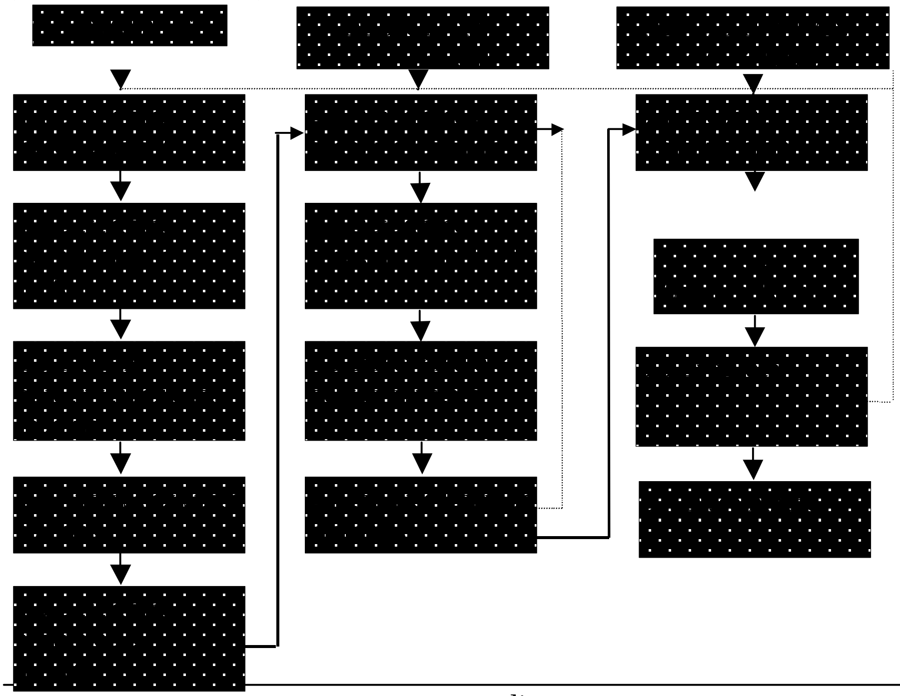
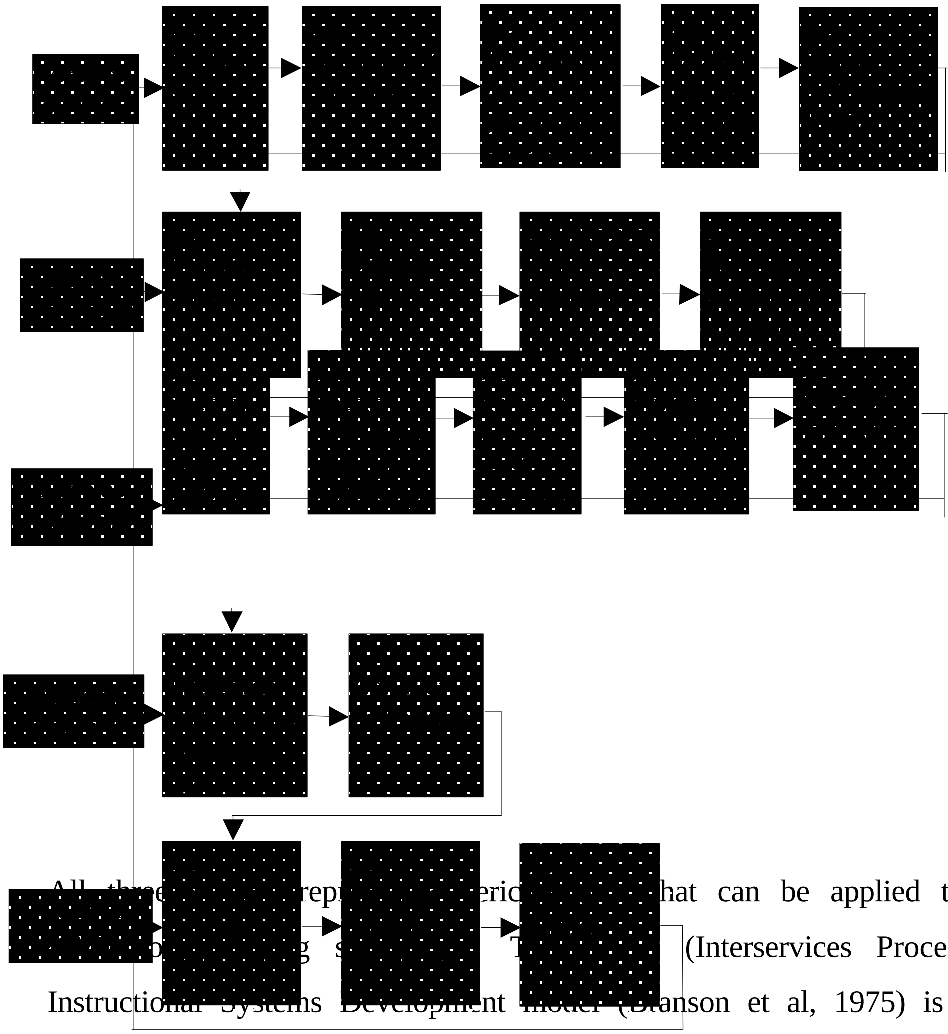


Figure 5.3: The Interservices Procedures for Instructional Systems Development (IPISD) model (Branson et al, 1975; 1977)



All three [illegible] rep[illegible] eric [illegible] hat can be applied to various [illegible] g s[illegible] T[illegible] (Interservices Procedures for Instructional Systems Development model (Branson et al, 1975) is the most commonly known. It was developed by the U.S. military to highlight the important functions when developing a training programme. All three models have generally the same stages of development (also referred to as phases, goals, functions, components or tasks in other models). Firstly the job or task goes through some analysis or needs assessment, then the training materials are designed and developed, the training is implemented, the outcomes evaluated and the results are fed back into the system. Patrick (1992) lists five advantages of ISD models as:

(1) *ISD models identify generalisable functions in the development of training.* The aim of these models is to provide a list of functions to be performed when developing any training programme, therefore they are intended to be generic or 'context independent'.

(2) *ISD models are helpful to those unfamiliar with training development.* As the models highlight the functions of the training development process, they are a useful initial guide for those unfamiliar with this process.

(3) *ISD models are particularly useful to large-scale organisations.* As the stages are divided into separate functions, it is easier to subdivide the work and schedule work as some functions have to be performed before others.

(4) *Psychological principles can be appended.* As discussed above (section 5.3.1) there are psychological theories that are relevant for all the stages of the training process which can provide useful insight when performing each function.

(5) *An ISD model can be used as an evaluation framework.* If when evaluating outcomes there appears to problems this could be due to the development process of the training. The ISD model can be used to examine each function of the development process to highlight what has caused this effect. For example, if the task was inadequately analysed, this effects every subsequent function and may result in an ineffective or even detrimental training application.

Patrick (1992) furthermore lists two disadvantages of ISD models.

(1) *It is an idealised top-down view of training development.*

The view presented of training is quite idealistic and disguises the fact that most tasks have their own individual considerations.

(2) *ISD models specify 'what to do' rather than 'how to do it'.* This is related to the previous point, as ISD models make the functions appear quite clear and straightforward however each function can be extremely difficult to achieve successfully. However the models are there to provide generic guidelines and should not be considered as the only technique required to develop successful training programmes.

The main stages of the development process will now be discussed within a systematic framework but with support from psychological theories.

5.4 Analysis of Training Needs

Analysis of training needs is one of the fundamental processes in training development as it forms the basis of all the other training activities in the development process (Boydell and Leary, 1996). It primarily involves three levels of analysis (Goldstein, 1993) - organisational (or business analysis); job analysis (task and knowledge, skill and ability analysis); and person or user analysis. These will be discussed as follows, along with some of the main techniques of investigation involved at each level.

5.4.1 Organisational Analysis

There are a number of factors that have an impact on organisational needs. These have been described by Bee and Bee (1994) citing the work of Robbins (1988) as: 'general external environmental factors' e.g. economic, political, social/demographic/cultural and technological; 'specific environmental factors' e.g. customers, suppliers, competitors and pressure groups; and 'the internal environment' e.g. employees, trade unions and shareholders. In light of these factors organisations are required to develop a strategy to deal with them. Such strategies include (Bee and Bee, 1994): (i) *consolidation* (strengthening the current position); (ii) *growth* (by market penetration, market development, product development, diversification, etc.); (iii) *contraction* (reducing activities by withdrawing from markets, withdrawing products, selling off or closing down parts of the business, etc.); or (iv) closure (dissolve the whole business). These strategies are important to recognise as they provide an opportunity for training, e.g. training may benefit consolidation by generating cost and time savings thus increasing efficiency; it may benefit growth, as changes in the current business may require new skills; and it may benefit contraction or closure by re-skilling the current workforce so that less new employees are required. Therefore, organisational analysis concerns *"examination of short-term and long-term goals* *of the organisation and trends that are likely to affect these goals"* (Goldstein, 1993). In general there are two aims of organisational analysis for training: firstly to identify a need for improvement in performance; and secondly to identify whether training would be the most appropriate solution. Other potential solutions to a need include: personnel selection or ergonomics or a combination of all three (Patrick, 1992). For example, one potential option is to recruit people who have the appropriate abilities, attitudes or previous training to deal with the problems of a particular job, alternatively ergonomics could be applied to re-design the task to make it easier to perform. Or the appropriate people could be selected, then trained to perform the ergonomically-designed task. The main techniques for identifying organisational needs is through discussions with top-level management and examination of company policies. Other techniques are described later in section 5.4.4.

5.4.2 Job analysis

The next level of analysis involves 'job analysis', this includes analysis of the functions, tasks, knowledge, skills and abilities (KSAs) of the job. Further to the investigation techniques described in section 5.4.4 the most popular technique used at this level is 'task analysis'. An overview of this method can be found in Stammers and Shepherd (1995), in particular, the most well known type is the Hierarchical Task Analysis (HTA) developed by Annett and Duncan (1967; cited by Patrick, 1992; Goldstein, 1993; and Stammers and Shepherd, 1995). This technique involves the tasks being broken down into increasingly specific operations in a hierarchical way. The result of a task analysis is a breakdown of the operations required to perform the job however this alone cannot provide all the necessary information required to design the training. Goldstein (1993) suggests that to provide training on the exact tasks that exist on the job would require the training system to have a very high 'physical fidelity', that is, a high "*representation of the real world of operational equipment*". However not only is this extremely difficult (and costly) to achieve (hence the high costs of simulators) but in some cases an exact representation may be too overwhelming for the trainee e.g. learning how to fly. Therefore Goldstein (1993) suggests that the aim of the training should be on 'psychological fidelity', that is, an exact "*representation of the essential behavioural processes necessary to perform the job*". Therefore it is necessary to analyse the relevant knowledge, skills and abilities (KSAs) required of the trainee to perform the task. Prien (1977; cited by Goldstein, 1993) defines the attributes as follows:

"Knowledge (K) is the foundation on which abilities and skills are built. Knowledge refers to an organised body of knowledge, usually of a factual or procedural nature, which if applied makes adequate performance possible. It should be noted that possession of knowledge does not ensure that it will be used.

Skill (S) refers to the capability to perform job operations with ease and precision. Most often skills refer to psychomotor-type activities. The specification of a skill usually implies a performance standard that is required for effective job operations.

Ability (A) usually refers to cognitive capabilities necessary to perform a job function. Most often abilities require the application of some knowledge base."

These KSAs should be linked to the task analysis in terms of what KSAs are required for each operation to be performed effectively. (There is however some confusion with this term as some authors refer to knowledge, skills and 'attitudes' and do not use the acronym KSAs.)

Patrick (1992) also provides a description of other approaches which are associated with this particular level. These include: Miller's 25 task functions (Miller, 1973; and summarised in Fleishman and Quaintance, 1984); Altman's motivational, behavioural and contextual domains (Altman, 1976); Position Analysis Questionnaire (PAQ) (McCormick et al, 1969, 1972); 'ability requirements' approaches (reviewed in Dunnette, 1976; Ekstrom, 1973; Fleishman and Quaintance, 1984; and Sternberg, 1985); and 'types of learning' (reviewed in Glaser and Bassok, 1989).

5.4.3 Person analysis

Finally person analysis or user analysis provides the answers to the questions of who needs training and what kind of training is needed (Goldstein, 1993). It also considers how well a person is carrying out the tasks involved in the job (McGhee and Thayer, 1961). As such this is linked to expressing the task in terms of KSAs (discussed above, section 5.4.2), except the focus is how well the person performs the KSAs required for the job. In general, the main methods employed for this analysis tend to be interviews and questionnaires.

5.4.4 Techniques for investigation

To identify these needs various techniques of investigation are used (for further information refer to Bee and Bee, 1994; Craig, 1995; and Boydell and Leary, 1996) . The main ones are discussed briefly as follows:

5.4.4.1 Brainstorming

This technique was developed by Alex Osborn (Osborn, 1941; cited by Craig, 1995) and involves either an unstructured or structured approach of a group of people generating ideas around a theme e.g. how can we increase productivity?, how can we decrease errors in performance? etc. Craig (1995) lists four rules to be observed:

"(1) Ideas are to be freely expressed without any questioning about feasibility or usefulness; anything and everything is accepted.

(2) Discussion, categorising and rejection/acceptance of ideas are done only when the actual brainstorming exercise is at the end and this end must be made clear to the group by the leader.

(3) Members are to be encouraged to build on ideas already expressed or to modify these ideas with further ideas.

(4) Quantity is asked for, the aim being to get as many ideas as possible." (p. 7)

The approach can be structured by having a leader of the session who controls the process. After the initial generation of ideas, these can be categorised into - good ideas, ideas to be considered and ideas to be rejected. The process ideally should have at least 45 minutes of uninterrupted time. An unstructured approach has no defined leader.

5.4.4.2 Critical Incidence Technique

This technique was developed by John Flanagan (1954; cited by Craig, 1995; Bee and Bee, 1994) and as its name suggests it involves identifying the critical influences on the success or failure of a job. The technique is usually carried out through interviews but questionnaires can also be used. A number of key incidents are identified to be investigated and the people involved in these incidents are asked to recall the events leading up to them. The advantage of this technique is that it is quite straightforward and requires little training of interviewers, however it relies on the memory of the individuals involved.

5.4.4.3 Repertory grid interviews

This is based on the Personal Construct Theory developed by Kelley in the 1950s and described in detail by Stewart, Stewart and Fonda (1981; cited by Bee and Bee, 1994; Craig, 1995). Kelley identifies personal constructs as a way of getting insights into people's view of the world, in particular views on what constitutes good and poor performance. The example provided by Bee and Bee (1994) involves a manager asked to select nine employees, three of which are considered as effective performers, three as ineffective performers and three, whose performance is variable. The names of the employees are placed on cards which are laid out in a 3x3 grid formation. Three cards are then chosen and the manager is first asked to match the two people who are the most similar and describe what their similarities and differences are. Then the manager must describe how the third person is different or similar to the two chosen. The result is a huge amount of information about the matching characteristics of effective performers and ineffective performers. The advantage of this technique is that it provides a structured way of identifying people's perceptions of good and bad performance. However it requires a lot of skill on the part of the interviewer to guide the interviewee.

5.4.4.4 Observation techniques

Direct observation of people in the work place is probably the most obvious approach to gathering information. The advantage is that direct evidence of the job performance can be noted, however interpretations of this are subjective and therefore require some skill. It is also very time-consuming. Observation can be carried out directly or indirectly (through video analysis after the performance) however it only provides information on the physical attributes of the job and not necessarily on the intellectual processes. There is also problems with the 'Hawthorne' effect, i.e. subjects under study change their behaviour because they are being studied (Dane, 1990).

5.4.4.5 Surveys, Questionnaires and Interviews

The main advantages of surveys and questionnaires is that they are usually quick to administer and can yield a large amount of data. The major disadvantage is that they need to be carefully designed in order to be effective. A good guide to the design of questionnaires is Oppenheim (1992) and an introduction is provided in Dane (1990) and Sinclair (1995) . They are a good source of subjective information and can be delivered either through interviews (face-to-face or telephone) or through less direct means (straight after the job performance or through the mail). Generally there are three types of questions that can be employed - structured, semi-structured and open. Structured questions involve some type of multiple-choice where a list of answers or a rating scale, are provided to the respondent, who has to choose the most appropriate answer. Semi-structured questions are those that have set questions followed by an opportunity to elaborate on their chosen answer. Open questions are questions which allow the respondent to provide further information. The actual design is dependent on the type of information that is required.

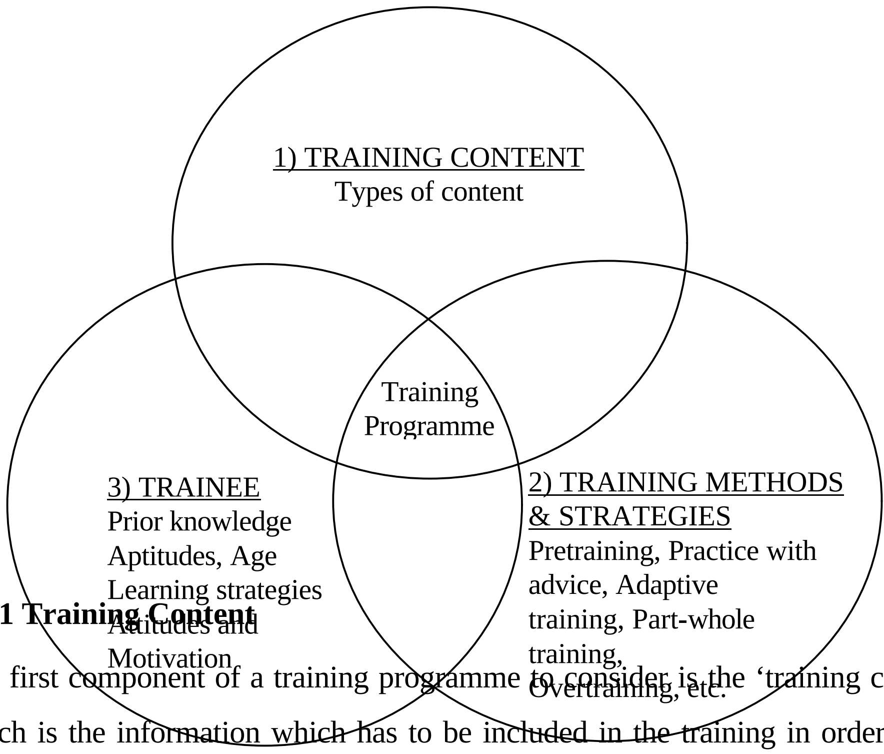
5.4.4.6 Delphi approach

This method was developed by the Rand Corporation in the U.S. and descriptions are provided in Linstone and Turoff (1977), Mullen (1983) and Foley (1984) (all cited by Craig, 1995). This method is a 'group information-collecting technique' through questionnaires. A group of experts are given a questionnaire and the opinions are distributed back to the group (preserving anonymity) in order that they can reconsider their opinions based on what others have said. The advantage of this method is that group opinions can be identified eliminating the influences of more outgoing/outspoken characters in the group. Also the group can involve people outside of the organisation e.g. customers, suppliers, etc. The major disadvantage is that it is necessary to be highly skilled in questionnaire design to employ this technique.

5.5 Design and Development of Training

Patrick (1992) suggests that there are three components of a fully designed training programme that need careful consideration. These components are illustrated in Figure 5.4. Each component not only has inputs into the training programme, but they also have effects on each other.

Figure 5.4: The three components involved in a training programme (based on Patrick, 1992; p.272)



5.5.1 Training Content

The first component of a training programme to consider is the 'training content', which is the information which has to be included in the training in order for the trainee to effectively learn the job. This should have already been defined at the training needs analysis stage (section 5.4).

5.5.2 Training methods and strategies

Another component to consider in the training programme is the method of training (also known as the delivery method) and the strategy of the training programme. There are various theories on learning and skill acquisition that exist in order to provide understanding of this area and to guide how the training is structured. Patrick (1992) and Goldstein (1993) discuss some of the better known theories, these are briefly listed as follows:

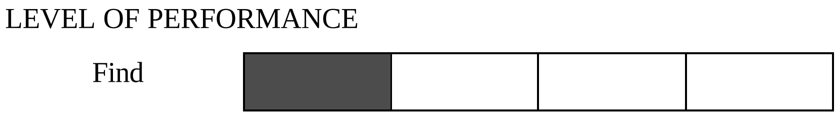
1) Gagné's theoretical formulations (Gagné 1977, 1985)

Gagné's theoretical formulations are probably the most well-known and influential in the training field. In particular, Gagné is attributed as one of the first to realise that all learning is not the same and learning outcomes require different learning conditions. Learning outcomes were divided into five categories: 1) intellectual skills (sometimes referred to as procedural knowledge) includes concepts, rules and procedures; 2) verbal information, sometimes referred to as 'declarative information' where a trainee can declare or state something; 3) cognitive strategies, are the knowledge that the trainee brings to a new task; 4) motor skills, are the skills that require physical motion like using tools etc.; and 5) attitudes, which are the trainees opinions about a subject.

2) Merrill's Component Display Theory (Merrill and Boutwell, 1973; Merrill et al, 1977, Merrill, 1983)

This is shown in Figure 5.5. Merrill proposes three levels of performance - remember, use and find - and four levels of content - fact, concept, procedure and principle.

Figure 5.5: Performance-content matrix (Merrill, 1983)



Use				
Remember				
	Fact	Concept	Procedure	Principle

TYPE OF CONTENT

The model allows a combination of ten types of learning to be identified (Merrill excludes the ‘use-fact’ and ‘find-fact’ components).

3) Landa’s Algo-heuristic theory of instruction (Landa, 1983; summarised by Reigeluth, 1987)

This theory generally highlights the importance of allowing the trainee to self-discover the rules and heuristics of skilled performance.

4) Rumelhart and Norman’s (1978) tri-modal taxonomy of learning.

This describes three types of learning - accretion, restructuring and tuning - defined as follows: “Accretion refers to the acquisition of facts in declarative memory. Restructuring refers to the initial acquisition of procedures in procedural memory. Tuning refers to the process of modifying existing procedures in procedural memory to make such procedures quicker and more reliable.” (Kyllonen and Alluisi, 1987; p.133, Patrick, 1992)

5) Fitts’ three-phase theory (Fitts, 1962)

This considers the three phases or stages of skill development: (1) the cognitive phase where the ‘intellectualisation’ process of learning a new task takes place, (2) the fixation or associative phase where correct patterns are established through practice with errors gradually diminishing; and (3) the autonomous phase where the skill becomes automatic.

6) Adams’ closed-loop theory of motor learning (Adams, 1971, 1987)

This considers two conditions that explain learning - a perceptual trace and memory trace. When performing a motor skill the movement generates a ‘perceptual’ trace which is eventually established as a memory trace after a number of successive repetitions.

7) Schmidt’s schema theory of motor learning (Schmidt, 1975)

This suggests that a schema or pattern of the learner's experiences is stored in the memory to guide future actions. This schema enables us to abstract or generalise to other situations.

8) MacKay's theory (MacKay, 1982)

This suggests a hierarchy of interconnecting nodes where at the top there are mental nodes or processes which feed down to the processes that control muscle movement. This system has three properties - activation, priming and linkage strength. All the nodes need to be activated in the correct order to provide the appropriate muscle movement. They then become primed so that when activated again, they can perform more quickly having stored the actions. Strengthening occurs with practice.

9) Anderson's theory of cognitive skill acquisition (Anderson, 1982, 1983, 1987)

This is also know as the ACT* model and is similar to Fitts' and to Rumelhart and Norman in that it also proposes three stages of skill development - declarative stage, knowledge compilation stage and tuning stage. At the end of the declarative stage, the trainee has knowledge about the task, by the end of the knowledge compilation stage the trainee gains procedural knowledge and by the end of the tuning stage the skill is firmly established in procedural memory.

There are many more such models described in Patrick (1992) and Goldstein (1993) however the ones listed above are among the most cited by other authors for their influence in the area. These models offer a strategy for design by providing understanding and recommending the stages that are required in a training programme in order to learn and reinforce various skills. Other issues in design including: pre-training, practice, adaptive training, part-whole training and over-training, etc. These will be determined by the training content and the existing capabilities of the trainee. These will also determine the eventual delivery method of training, as the methods must be able to convey the training content effectively and promote the intended change in the performance of the trainee by using their existing knowledge, skills and abilities. The many types of delivery methods currently available are wide and varied and are beyond the scope of this thesis, however detail specifically about computer use in training is considered at the end of this chapter (section 5.7).

5.5.3 The Trainees

The final component to consider in the training programme is the trainees or target audience. This includes issues like prior knowledge, experience, age, learning strategies, attitudes and motivation (Patrick, 1992). In particular, prior knowledge and experience are necessary information in order to be able to set the level of training required e.g. if the trainees are novices then more training is required than if the trainees are semi-skilled. Also this information can provide the baseline for evaluations, as any new knowledge, skills or abilities (KSAs) and attitudes gained after training compared to the trainee's initial state may provide evidence that the training was effective. Age may also have implications, Patrick (1992; chapter 10) cites a number of studies performed to look at various differences in performance between people of different ages. Also trainees may have different learning strategies (some were mentioned in section 5.5.2), attitudes and motivations which can be used to make the training more engaging (and can also explain why the training was ineffective e.g. negative attitude towards computers, etc.).

5.5.4 Training Specification

The consideration of the above components of the training programme can be outlined in a training specification. This is defined as "*a blueprint or detailed plan for the training required to meet the gap in performance and for measuring its effectiveness*." (Bee and Bee, 1994; p. 108) Much of this specification will come from the training needs analysis carried out at the beginning of the development process. Structures of these specifications may vary but they should at least include the following information (Bee and Bee, 1994):

5.5.4.1 Background to the organisational need.

This information should come from the organisational analysis (see section 5.4.1), it can also be thought of as the 'definition of the problem and is important as it gives an indication of the driving force behind the training programme. For example, the problem could be that the present training method is uneconomical and there is a need to reduce costs. Therefore the driving force behind the new training method will be to ensure that the new method chosen is more economical than the current method.

5.5.4.2 Description of the target training population.

This information should come from the person analysis (section 5.4.3) which describes the necessary characteristics (e.g. age, gender, educational profile, previous training) and knowledge, skills and abilities (KSAs) and attitudes required to perform the task. Also the number of expected trainees and their possible locations should be defined.

5.5.4.3 Overall aim of training.

This is linked to the organisational need and not only does it provide the focus for the training objectives but also it provides criteria or benchmarks for the evaluation process. For example, if the reason for the training is to increase efficiency of production through time savings one of the solutions could be training to reduce errors by 50%. Therefore the type of training required should be focused on reduction of errors and the criteria of success would be evidence that errors have been reduced in the work situation by 50% and this can be linked to an increase in productivity.

5.4.4.4 The training objectives.

By far the greatest influence in this area is Mager (1962, 1991; cited by Patrick, 1992; Goldstein, 1993; Sanderson, 1995; Bee and Bee, 1994). Mager (1991) describes the characteristics of objectives as three components - performance, conditions and criterion. 'Performance' describes what the trainee is expected to be able to do by the end of the training or what Mager calls the 'terminal behaviour' e.g. to reduce the number of performance errors, to operate a machine, to recognise certain parts, etc. 'Conditions' describes under which conditions or limitations the performance is to occur e.g. given the standard tools, given two hours, etc. Finally, 'criterion' describes the acceptable performance e.g. by 50%, within one hour, according to company policy, etc. However even though many authors cite Mager's three components as a good way of detailing objectives, they also state the difficulty in producing objectives in this way. Sanderson (1995) comments on an educational programme that had 10,000 objectives and also cites the work of Davies (1976) which showed that specific behavioural objectives were not more effective than general objectives. Other approaches include just describing the intended behaviour or using measures that already exist e.g. National Standards Programme (Training and Development Lead Body, 1991)

5.5.4.5 The training methods to be used.

This has already been discussed in section 5.5.2. The training method should be based on the training content, the structure of the delivery and the characteristics of the trainees (as discussed in section 5.5.2)

5.5.4.6 Skills required by the trainer(s).

Training using different types of methods require different levels of expertise. It is important to list the precise skills and qualifications required, as well as the level of experience so that appropriate trainers can be found. For example, if the training method uses computers than the trainers should have some knowledge of computers or at least the training package used.

5.5.4.7 How the training is to be evaluated.

The evaluation methods should ideally be considered at the beginning of the training although this is often missed out for two main reasons - firstly, many organisations do not evaluate their training either at all, or in a systematic way and secondly, many organisations do not even think about evaluation until the training has been delivered (Bee and Bee, 1994). However evaluation is very important for a number of reasons (discussed later in section 5.6) and it is easier to perform the earlier it is considered.

5.5.4.8 The time-scale for the delivery of training.

The time-scale of the training will be determined by: the urgency of the organisational need; practical constraints such as availability of trainers, facilities, trainees etc.; and budgetary constraints (Bee and Bee, 1994).

5.5.4.9 The learning environment.

If it is possible to make a choice, the specification should include the preferred learning environment e.g. classroom, training centre, home etc. Also the required equipment and facilities to fulfil the requirements of the trainee and the training programme e.g. audio-visual equipment, administrative aid, parking etc. (This is discussed later in section 5.6.3.4).

5.5.4.10 Any other constraints.

Finally, there may be other constraints which have to be considered which are particular to an application. For example, the amount of available facilities may limit the number of trainees that can be trained at the same time. Also the delivery method itself may pose limits on the amount of time the trainees are able to be trained. For example, in the case of simulators, some trainees are prone to symptoms of simulator sickness therefore there is a limit on the time the trainees are allowed in the simulators.

Given the identification of the training specification the training programme can then be designed and implemented.

5.6 Evaluation of Training

It should be possible to see from the previous sections that the evaluation of training is very much linked to the development process of training. This is because evaluation is ultimately concerned with measuring how far the training has achieved the objectives defined in the specification (discussed in 5.5.4) in terms of efficiency and effectiveness (Sanderson, 1995). This will be discussed next. However an important thing to remember is that the evaluation process itself has costs and benefits associated with it. Some of these have been listed by Breakwell & Millward (1995) and shown in Table 5.1. In terms of costs, the evaluation process will involve some financial outlay to pay evaluators, to design, produce and administer evaluation materials. This incurs some opportunity cost, that is, the cost of performing this activity is preventing the funding of another activity.

Table 5.1: The cost and benefits of performing an evaluation

Costs	**Benefits**
• entails financial outlay • incurs opportunity penalties • arouses distrust and anxiety • might yield unanticipated and unwelcome results • makes you aware of problems that are impossible to solve • provides ammunition for enemies • heightens internal competition • instigates external and internal opposition • initiates unrealistic expectations about change • involves penalties if findings are ignored or misconstrued	• estimates strengths, weaknesses opportunities and constraints • identifies problems and suggests solutions to them • specifies where you are succeeding • aids quality assurance • makes it evident that you are taking accountability seriously • improves credibility of basis for case for extra resources • heightens staff motivation if sometimes only temporarily • justifies change already planned • allows change to be monitored • creates standards for the future

Evaluations can also cause distrust and anxiety as trainees and trainers may feel that the results could lead to some negative outcome like job losses, which may make them uncooperative or bias the results of the evaluation (referred to as threats to internal and external validity, discussed in section 5.6.6.1). This is especially if the evaluations expose some unanticipated or unwelcome results, such as the inadequacies of the trainers or trainees; or even highlight problems that have no or impossible solution. Highlighting weaknesses can make the organisation vulnerable to competitors internally (between departments) and externally (between companies). Finally it may provide a false expectation of change which may lead to disharmony in the workforce if change is not implemented.

In terms of the benefits, evaluation can highlight the strengths, weaknesses, opportunities and threats to the organisation so that potential problems can be identified and potential solutions considered. It also highlights where the major strengths are so that these can be maintained. It ensures that each process and function is performing to capacity and indicates to the rest of the organisation that the management 'cares', thus (temporarily) increasing motivation. Finally evaluation can justify why a change had to be made, it can monitor this change and provide standards for future projects to be measured against.

5.6.1 Purposes of evaluating training

Many of the costs and benefits mentioned before provide an insight into the main purposes of evaluation. These have been grouped by Bee and Bee (1994) (and can be related to the definition of evaluating training provided by Goldstein, 1993 (section 5.2): where evaluation helps to make informed decisions about the "*selection, adoption, value and modification*" of training). The groups are as follows:

1) To improve the quality of training - this is also called formative evaluation (discussed later in section 5.6.10), the training method may be an old approach that needs modification or a new approach that needs assessing. Therefore the evaluation is particularly aimed at the content and design of the training.

2) To assess the effectiveness of training - this is concerned with whether or not the training has met the requirements of the training objectives defined in the specification. These objectives will determine the level of evaluation required.

3) To justify the training programme - this is concerned with providing information that will justify the current design of the training. This considers 'cost-effectiveness' information.

4) To justify the role of training - this is concerned with providing information that will justify the choice of training to meet organisational needs compared to other possible solutions. In particular, this requires some cost/benefit analysis and return-on-investment (ROI) information.

5.6.2 Levels of evaluation

The levels of evaluating training have traditionally been based on a number of models developed by Kirkpatrick (1967), Warr, Bird and Rackham (1970, 1978) and Hamblin (1974). Their models of evaluation consist of generally four levels (although Hamblin divides the fourth level into two, distinguishing between organisational objectives and ultimate economic value) and are often refered to as the four levels of evaluation. The corresponding parts of each model are shown in Table 5.2 below.

Table 5.2: Comparison of three of the main models showing the levels of training evaluation

Kirkpatrick (1967)	Warr et al (1978)	Hamblin (1974)
1. Reactions	1. Reactions	1. Reactions
2. Learning	2. Immediate Outcomes	2. Learning
3. Job Behaviour	3. Intermediate Outcomes	3. Job Behaviour
4. Results	4. Ultimate Outcomes	4. Organisation 5. Ultimate Value

Warr et al, (1978) however provides a much broader view of training by including two pre-ceeding levels (specifically concerning the development process of training) before the ones shown in Table 5.2. These levels are ‘context’ and ‘input’ evaluations, resulting in a framework called C.I.R.O. - Context, Input, Reaction, Outcomes (at three levels). A discussion of each area follows.

5.6.3 Context evaluation

Context evaluation considers gathering information about the context of job or task, in order to determine the training needs and objectives. This corresponds to the training needs analysis stage which has already been discussed in section 5.4. Sanderson (1995; p. 127) lists some of the key questions in this type of evaluation (adapted from the CIRO framework) as: *what needs to be changed? Is the* *training solution appropriate? Are the objectives the right ones? Do they relate to the training needs analysis? Is the training needs analysis acceptable? Are the objectives clear, achievable and measurable? How will we measure immediate, intermediate and ultimate objectives?*

5.6.4 Input evaluation

Input evaluation considers gathering information about the available training resources in order to make decisions about the inputs to the training programme. This corresponds to the design and development stage which has already been discussed in section 5.5. Again, Sanderson (1995) lists some of the key questions in this type of evaluation (adapted from the CIRO framework) as: *what procedures are most likely to bring about change? How much time is available? What are the relative merits of different training methods? What were the results of previous similar courses? Should you use an external training organisation? What should be the content and what evidence is there to support the choices made? Does the content reflect the objectives?*

5.6.5 Reaction evaluation

Reaction evaluation is concerned with the attitudes and opinions of the trainees towards various features of the training depending on the purpose of the evaluation. These include: "*pre-course briefing or joining instructions, objectives, content, methods, resources, facilities/accommodation, duration, relevance to job/intended changes, tutor, general comments etc.* (Sanderson, 1995; p.130). A questionnaire using rating scales is usually the most popular method of assessment and this can be supported by interviews and informal discussions. Sanderson (1995; p. 127) lists some of the key questions in this type of evaluation (adapted from the CIRO framework) as: *what are the trainees' opinions of the training? Should there be an end-of-course paper and pencil review or follow-up? Should these be anonymous? Should there be evaluation of each session as well as at the end of the course? Should you use a rating scale or questionnaire? Should you convert answers to numerical score to make comparisons possible? What should* *the content of the review be? Should it be conducted by the trainer or a neutral observer? Should session reactions be fed back to the tutor to improve the on-going event?*

5.6.6 Immediate or learning outcomes

This level is concerned with measuring the changes in the trainee that are the result of training. These changes should have been defined as objectives at the beginning of the development process in the specification (section 5.4.4.4) and the design and development process of the training should have ensured that these objectives can be achieved. However if a change is failed to be shown, this could still be due to the training programme rather than the trainee or even in the actual methods used for the evaluation study.

5.6.6.1 Validity and Reliability of outcomes

There are three aspects that need to be considered - internal validity, external validity and reliability. Generally, internal validity refers to how well the study measures what it is designed to measure, that is, when a change has occurred can the study easily identify this. A number of threats to internal validity have been listed by Goldstein, 1993 and Patrick, 1992 (citing Bracht and Glass, 1968; Campbell and Stanley, 1966; Cook and Campbell, 1976) and Dane, 1990 (citing Campbell and Stanley, 1963) as follows:

1) History - specific events, other than the training, that occur between the first and second measurement, e.g. a job accident in the company between two training sessions may reveal a change in attitude towards training between the first and second session.

2) *Maturation* - biological or psychological effects that systematically vary with time e.g. participants become older or less interested in the training over a period of time.

3) *Testing* - the influence of the pre-test on the scores of the post-test

4) *Instrumentation* - changes in the measurement instruments pre- and post-test

5) *Statistical regression* - the phenomenon which occurs when on second testing of the scores of two extreme groups, the scores regress towards the middle of the distribution

6) *Differential selection of participants* - evaluator's biases when selecting participants for groups

7) *Experimental mortality* - the loss of participants during the study

8) *Interactions* - the interactions of any of the factors (1 - 7) listed above, e.g. the study of a group of younger people and older people over a period of time, the results may be effected by differential selection and maturation.

9) *Diffusion or imitation of treatments* - this refers to the participants passing on information to each other before the study

10) *Compensatory equalisation of treatments* - this refers to the different groups ending up equal in some way because other factors were introduced to compensate them for being in different groups.

11) *Compensatory rivalry between respondents receiving less desirable treatments or resentful demoralisation of respondents receiving less desirable treatments* - this refers to the participants being aware of which group they are in, which may result them changing their behaviour.

'External validity' refers to the "*generalisability of the study or the extent which the results of the evaluation can be applied beyond the groups chosen in the study*" (Cook, Campbell and Peracchio, 1990; cited by Goldstein, 1993). The following are threats to external validity Goldstein (1993):

1) *Reactive effect of pretesting* - the effects of pretesting may lead to the participants paying more interest to the training, however when the training is carried out in the working environment the effects may not be the same.

2) *Interaction of selection and experimental treatment* - the characteristics of the group selected for evaluation of the training may not sufficiently represent the eventual trainees who will be using the training.

3) *Reactive effects of experimental settings* - this is also refered to as the 'Hawthorne effect', where participant's under study change their behaviour because they are being studied so that the results cannot be generalised to the eventual trainees who will not be studied.

4) *Multiple-treatment interferences* - the effects of previous studies still having an effect on the participant.

Finally, 'reliability' is the extent to which the results are consistent in numerous studies. Bramley (1996) suggests that about 20% of the results of a sample of participants are likely to be unreliable because some people have a better aptitude and attitude to test measures (e.g. exams, questionnaires, etc.), some people will tend to perform better on some days and at different times of the day. Also some errors will be made in the scoring. However various statistical tests are designed to eliminate these likely chance effects in the results. Validity and reliability vary according to the actual design of the study.

5.6.6.2 Experimental Design

In order to measure change, some form of experiment is usually designed. There are many different types of designs of experiments which depend on the purpose of the evaluation study. Many of these are discussed in Dane (1990; chapter five and six) and Goldstein (1993; chapter six). The main ones are illustrated in Table 5.3 (overleaf) based on these texts.

Experimental design is concerned with the "*number and arrangement of the independent variable*" (Dane, 1990). The 'independent variable' is the "*expected cause under investigation"* in this case it is training.

Table 5.3: Examples of the main experimental designs (based on Dane, 1990; and Goldstein, 1993)

Design	**Group**	**Pretest**	**Training**	**Posttest**
Posttest-only	training		x	x

Pretest/Posttest	training	x	x	x
Pretest/Posttest	control	x		x
control-group	training	x	x	x
Solomon four-group design	control(1)	x		x
	training(1)	x	x	x
	control(2)			x
	training(2)		x	x

The 'dependent variable' is the "*effect under investigation*" or the learning outcomes which can be expressed by the performance of the participant. This is because learning cannot be directly measured, it can only be inferred by other measures like performance (Patrick, 1992). Also the groups may have a 'between-subjects' (different subjects in each group) or 'within-subjects' (the same subjects in each group) design. The former however may suffer from the groups being unequally matched and the latter design may not be possible to do.

The simplest experimental design involves just the training group(s). A 'posttest-only' design is the most basic, where the participants receive no pretest but rather are trained and then measured with a post-test. However it is then difficult to distinguish the changes just due to training from the internal threats to validity (discussed above section 5.6.6.1). However it is easy to carry out and can produce a lot of information. Alternatively a 'pretest/posttest' design can highlight a change from the initial state of the participant, measured by a pretest, to the final state of the participant, measured by a posttest. However it is still difficult to tell whether these changes have just occurred by chance or whether they are due to the training. Therefore a more effective design involves a control group also called a 'pretest/posttest control-group' design. This is where the control group carry out the pretest and posttest and any effects from this can be eliminated from the results of the training group(s) to leave just the effects of the training. Another design which considers the threats to external validity (discussed in section 5.6.6.1) is the

‘Solomon four-group’ design which adds a further two groups that are not pretested in order to eliminate the effects of pretesting on the training results.

Quasi-experimental designs include “*research methods that approximate but are not truly experimental methods*” Dane (1990). One such method is a time-series design where the participant repeatedly completes the pretest over a time period and at certain intervals, then is trained, then repeatedly completes the posttest for another length of time. This method controls time and maturation threats to internal validity but cannot control for effects of the pretest on the training. Also a ‘non-equivalent control-group’ design where the participants are not able to be randomly assigned to the groups but are pre-determined. This is still an acceptable design as long as the differences are considered along with the final results. Sanderson (1995; p.127) lists some of the key questions in this type of evaluation (adapted from the CIRO framework) as: *what changes in the knowledge, skills and abilities have resulted (or have the immediate objectives been met)? How can we measure changes?*

5.6.7 Intermediate or job behaviour outcomes

Intermediate or job behaviour evaluation measures the effect of the training on the eventual job performance. This examines whether the learning that has taken place and identified at the immediate level stage has been transferred to the workplace - this should be the main reason for training. The techniques which are usually employed for this method include observation, questionnaires and interviews (discussed in section 5.4.1) and experiments on transfer-of-training. The main issues of this have been discussed by Holding (1987; pp. 955 - 958); Patrick (1992; pp. 75 - 107) and Wickens, 1992; pp. 237 - 238). The basic measure of transfer-of-training is a percent and the simplest design involves a training group and a control group. The training group receive some training and then performs the real task and the control group just performs the real task. If the training group perform the task better than the control group, then this shows ‘positive’ transfer, if the control group perform the task better than the training group, then this shows ‘negative transfer and if there is no difference between the groups than this is ‘zero’ transfer (Goldstein, 1993).

The following equation can be used to calculate the savings in percent of the transfer-of-training (Patrick, 1992):

$$\% \text{ savings} = \frac{B_2 - B_1}{B_2} \times 100$$

where B_1 = number of training sessions or hours taken to learn task B after training on task A (or training group)

B_2 = number of training sessions or hours to learn task B with no prior training on task A (or control group)

Sanderson (1995; p.128) lists some of the key questions in this type of evaluation (adapted from the CIRO framework) as: *what changes are there in on-the-job performance (or have intermediate objectives been met)? How can we be sure that these changes are the result of the training?*

5.6.8 Ultimate or organisational outcomes or results

Ultimate or organisational outcome evaluation is concerned with the effects of the training on the organisation and should be related to the results of the organisational analysis. Bee and Bee (1994) suggest five stages to the process as follows: identify the key indicator(s)/measures of organisational performance e.g. safety training, customer care, etc.; ensure that the results are in an appropriate form in order to measure their initial state before training; decide how long the training will take to affect these indicators; identify and consider how to minimise factors that may intervene; and set-up systems to monitor the results. Even though this evaluation would provide the final answer to the question of whether the organisational need was fulfilled by the training, it is very rarely performed (Bee and Bee, 1994). Due to this the National Training Task Force developed the Investors In People (IIP) standards in order to guide the linking of training to organisational success, this is discussed at the end of the chapter (section 5.8). Sanderson (1995; p.128) lists some of the key questions in this type of evaluation (adapted from the CIRO framework) as: *how have the changes in job performance affected the organisation (or have the ultimate objectives been met)? What overall effect on profitability or effectiveness is there?*

5.6.9 Ultimate Value or Cost effectiveness and Cost/benefit of training

Finally the ultimate value of training to organisations is judged mainly in terms of its financial effect. Generally the organisation is concerned with 'cost-effectiveness' analysis i.e. the estimation of training costs and benefits in monetary terms (Sanderson, 1995) and 'cost benefit analysis' i.e. the examination of training in monetary terms compared to the benefits expressed in non-financial terms - improved attitudes, morale etc. (Sanderson, 1995). Methods for calculating costs can be found in Sanderson (1995) and include: Human Resource Value (HRV), Difficulty- Importance-Frequency (DIF) analysis (Cascio and Ramos, 1986) and Payback time.

In theory, if the training development process has been carefully designed according to the stages mentioned before, then the training applied should have been the most cost-effective option. However the best way to assess this is by measuring the training against a different approach in terms of which best satisfies the training objectives (Bee and Bee, 1994). Generally though cost-effectiveness studies of training are few in number and poor in quality (Patrick, 1992).

With regard to cost/benefit, this is usually the information of most interest but also the most difficult to provide. This is because it may not be straightforward to quantify benefits. One possibility is to use 'performance indicators' e.g. increase in sales, increase in productivity/output, reduction in wastage, accidents, equipment downtime, absence rates, turnover, labour costs, etc., (Bee and Bee, 1994). However it is sometimes impossible to attribute these factors solely to training. Craig (1995) offers five steps to conducting a cost benefit analysis which are summarised as follows:

1) State clearly the problem area e.g. this new training method is quite costly, what would be the costs and benefits of removing this training?
2) Collect various assumptions that can be made about the problem area e.g. the operators are having problems understanding how to use the new equipment so training is needed, the new equipment is very expensive and very easy to break, etc.
3) Identify cost and benefits of the training in consideration of the following: who?, when?, where?, why?, how?, what cost?, for how long? and how do we measure? - in financial and non-financial terms e.g. increased confidence, contentment, etc:
4) Compare the information. The financial costs and benefits can be used to calculate return-on-investment (ROI) which is usually the figure that the organisation is interested in. However it is possible for the ROI to show a break-even or negative result and the non-financial or subjective measures to show positive results (or the reverse).
5) Present the results. This includes the information from the previous four steps as well as an analysis of the outcome of the information in terms of recommendations.

5.6.10 Formative and Summative evaluation

The four levels of evaluation shown in Table 5.2 are generally conducted after training and are termed as ‘summative’ evaluation methods. Campbell (1988; cited by Goldstein, 1993) divides this into two types - ‘summative evaluation’ which refers to whether the training produced the expected outcomes and ‘comparative summative’ evaluation which refers to examining which of two or more training methods produced the better results. A second type of evaluation is ‘formative’ evaluation (attributed to Scriven (1967); cited by Tessmer, 1991; Goldstein, 1992; Patrick 1992; and Giber, 1997). ‘Formative’ evaluation is used to identify the strengths and weaknesses of the design of the training programme before implementation. Its major goal is therefore to improve the effectiveness of the instruction (Tessmer, 1991).

Before conducting any type of evaluation (formative or summative) there are a number of basic questions about the purpose and context of the evaluation which need to be answered. Many of the answers to these questions should come from the training specification outlined in section 5.5.4, these questions (formulated by Tessmer, 1991) include the following:

5.6.10.1 What are the goals of the evaluation

For summative evaluation these goals have already been discussed (sections 5.6.1 and 5.6.2) however the major goal of formative evaluations, is to improve the effectiveness of the instruction. This is achieved by four subgoals (Tessmer, 1991): (1) *identify deficiencies in learning effectiveness* - perhaps through a test, survey, interview or from on the job measures and deals with information like content quality, interest and motivation. The main question to answer is where in the training programme did the trainee fail to learn; (2) *locate possible usability problems* - including problems for the trainers, as well as the trainees e.g. is the training programme easy for the trainers to set up? will they need any special training? etc.; (3) *evaluate the efficiency of the instruction* - this is related to the time required for trainees to learn the training objectives and also the time taken for the trainer to apply the training programme. If learning time exceeds the amount of time available for the training then the programme fails to be effective; and (4) *analyse instructional strengths* - this is important for proving that the instruction is useful, for developing future training programmes and for excluding the successful areas from the areas that need to be improved.

5.6.10.2 Who is involved in the evaluation

It is important to know who is involved in the evaluation process as different people require different information e.g. the Finance manager wants to know about cost savings, the Production manager wants to know about time savings etc.

5.6.10.3 What outcomes will be assessed?

For summative evaluations these were discussed in section 5.6.2, however the major goal of formative evaluation is to revise instruction, so information just about the learning outcomes of the instruction does not provide information about how they did (or did not) learn. In order to find out this information it is necessary to consider the training objectives - what they were meant to learn. It also helps to know what the trainees knew before they began the training i.e. the entry skills. These should have already been defined at the beginning of the development process after the training needs analysis stage (see section 5.3). The different levels of analysis can be used to identify any difficulties. For example, the task analysis can be used to identify any steps that are being missed out and the person analysis can be used to highlight deficiencies in skills. (This is why it is important to ensure the analysis stage is accurate and considered carefully.)

5.6.10.4 What is the setting for learning

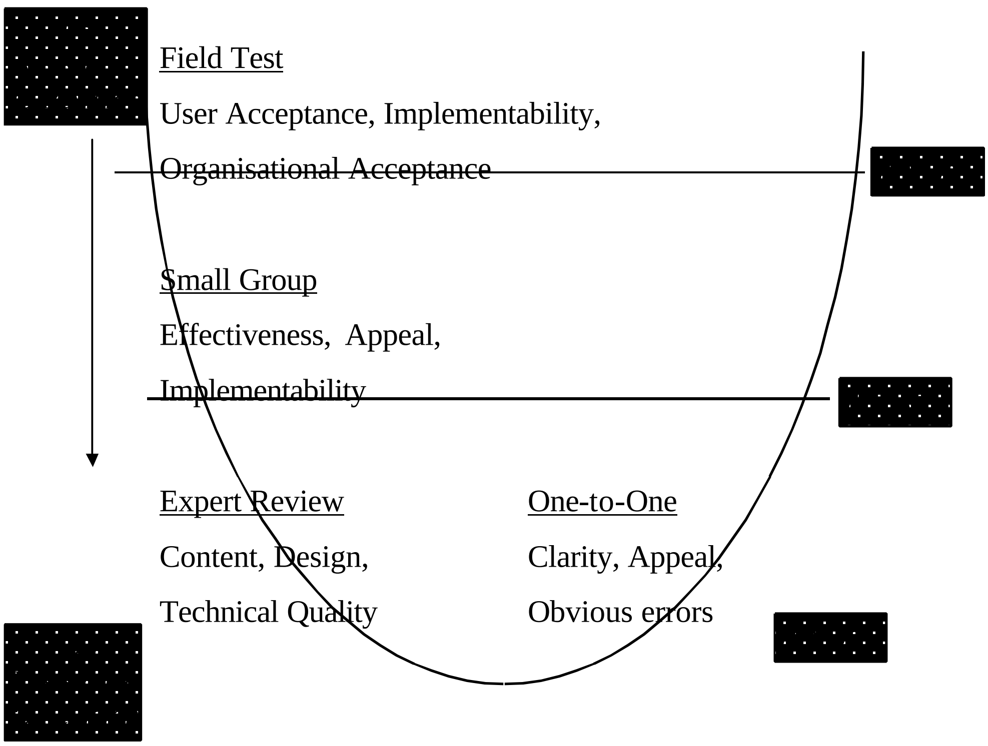
Tessmer (1991) describes this as environmental analysis, i.e. "*an analysis of the learning and support environments that 'surround' the instruction*". The first issue to consider is the requirements of the trainee in terms like time, equipment, materials or trainer. Secondly, there is the attributes of the learning environment which could be a classroom, training centre, computer laboratory, office or home or a combination of these. The environment is important as its physical characteristics may effect performance of the training e.g. lighting, seating, room size, etc. This also aids the evaluation methods which are set in experimental labs as the learning environments can be replicated to test for a variety of features, especially the effects of changing one feature of the physical characteristics e.g. different seating arrangements. Thirdly, the pattern of use of the training method may be important e.g. a training manual or computer-based-training (CBT) package may not be used in chronological order therefore it is important that it is structured so that information is found quickly and easily. Fourthly, it is important to assess the support mechanisms of the environment e.g. administrative or help services, technical support, maintenance and repair, etc. Also the availability of audio-visual equipment to be used e.g. overhead projectors, projection screens, videos, computers, etc. Furthermore the media intended to be used should be reviewed for their particular attributes e.g. 'navigation' is a distinctive attribute of text and CBT because it allows the user to go where they want whereas video does not. The evaluation should determine whether these attributes facilitate or hinder learning objectives

5.6.10.5 What are the available resources and constraints to evaluation

Different formative evaluation stages require different amounts of time, money and expertise. These different layers are shown in Figure 5.8 (overleaf).

The first layer of evaluation usually involves some form of 'self-evaluation'. Although this stage may seem obvious and usually a developer would automatically carry out a self-evaluation, it can be made more productive if a set of specific questions are used and possibly after a period of 'emotional detachment' (Thiagarajan, 1991). These questions can be devised from the task analysis and person analysis which should contain all the steps of the task and how they should be performed. The questions can also be used for the next layer of evaluation - expert reviews and one-to-one reviews.

Figure 5.6: Layers of Formative Evaluation (Tessmer, 1991; p. 35)



Self-Evaluation

Obvious errors

An expert review generally involves an expert (or different types of experts e.g. design, technology, subject matter, etc.) reviewing early versions of the instruction to determine strengths and weaknesses in terms of ‘content accuracy’ and ‘technical quality’ (depending on the type of expert used). Studies by Davidore and Reiser (1991) indicated that expert reviews can improve the learning effectiveness of materials revised with this procedure. They used a panel review by teachers to revise instructional material and students who used the revised version averaged 15.6 gain in points in post-test performance against those who used the unrevised material. Expert reviews can be asked at any stage of the evaluation process and the advantage over the other methods is that the experts have the knowledge and experience to provide informed feedback, whereas the other methods rely on the trainee. This can also be a disadvantage as the experts may not be aware of what the trainees requirements are. Generally this type of method needs careful consideration of two questions - what information is required? and which experts can provide this information?

A one-to-one review (also referred to as developmental testing) involves one trainee and the evaluator. This method can be used to provide information about the main characteristics of the training programme e.g. obvious errors, clarity of visuals, appeal, ease of use, etc. The main advantages are the information is provided from a potential trainee’s point-of-view and lots of specific information can be gained. Studies have been performed by Robeck, 1965; Wager, 1983; and Lowe et al, 1983 (cited by Tessmer, 1991) demonstrating the effectiveness of revised material produced by one or two learners. However a disadvantage is that the information is a limited view point which may or may not be reliable.

Small group evaluations usually involve a small group of trainees in an experimental setting and finally testing in the 'field' involves evaluation of the training in the actual training environment. These approaches are similar to those discussed in the immediate and intermediate levels of evaluation but the assessment is of the effectiveness of the design of the training based on the result of the outcomes (sections 5.6.6 and 5.6.7).

Knowledge of the resources and constraints can determine which of the layers shown in Figure 5.8 may be the most feasible. Ideally it would be beneficial to carry out all the layers of evaluation, however the first constraint is usually when the deadline for the project is, as the closer a project is to the end, the more resistance to modifications is met. Also time and money usually determines the types and numbers of evaluations that can be conducted. Generally expert reviews, one-to-one and small group evaluations are less expensive than field tests. Costs include personnel to conduct the evaluation and analyse the data although the main costs are the experts and non-experts. Time determines which evaluations are possible to complete and also what modifications are able to be performed before the intended implementation date.

5.6.10.6 What should be the measure of success.

There are many different measures of success that could be used to assess the application as suggested by the models discussed in section 5.6.2. For formative evaluations it is necessary to decide which information is the most important - the learning outcomes or the transfer to the work situation.

Therefore evaluation of every stage of the training development process is possible to be performed and the usefulness of the results are very much dependent on the early stages of the process where the training needs analysis defines the problems to be solved by the training. However even though there are many theories and methods and even systematic guides available to aid the process of evaluation, still very little is actually carried out. This will be discussed at the end of this chapter in section 5.8. However also relevant to this research is the particular use of computers in training. This is discussed next.

5.7 Computers and training

Patrick (1992) describes four main roles for computers in training: (1) provision of training - computers are obviously able to deliver training e.g. computer-based-training (CBT) packages, multimedia packages, simulation, etc.; (2) development of training - this involves the stages of development outlined before - needs analysis, design and development of training and evaluation - the computer can be used for any of these stages either directly or by supporting a person to do them. e.g. the design and production of training materials through word processing packages and art packages; (3) management of training - this is also referred to as 'computer-managed training' or 'computer-managed learning' (CML) where the computer is used for scheduling training, recording results, administering tests and producing reports; and finally (4) support of research into training e.g. research packages have been developed for studying workload, decision-making, stress and performance assessment, etc.

This section is particularly concerned with the provision of training i.e. the different computer methods available to deliver training. The use of computers can offer many benefits to an application but it cannot guarantee that the quality of training will be good (Patrick, 1992). What is required is careful consideration of the training development process as outlined in the sections before to ensure that training is effective and efficient. There is also the consideration of when it is appropriate to use computers. Seltzer (1971; cited by Patrick, 1992) suggested three possible conditions:

"*1) If the computer poses a unique solution to an important problem in the instructional process, then it should be used regardless of the cost involved.*

b) If the computer is more efficient or effective and the cost of its use to instruct is minimal, then it should be used ...
c) If the cost of development and use of the computer in instruction is relatively high with the relative efficiency or effectiveness only marginal, then the computer should not be used in the instructional process (p. 375)"

Holding (1987), Patrick (1992) and Goldstein (1993) all agree that the use of computers as training devices began with 'programmed instruction' (PI) or 'programmed learning' in the 1950s with the production of the 'teaching machine'. There were two major influential theories at this time, Skinner's theory of reinforcement (1954) and Crowder's (1960) intrinsic or branching programs. Skinner's theories resulted in the design of 'linear teaching programmes'. Generally these emphasised the correct way of learning something thus reinforcing the positive aspects and minimising learning of negative aspects of a task. The main characteristics of linear programmes are as follows: they are divided up into small steps or frames which provide information to the trainee; the trainee is expected to make a response and immediately receives feedback on whether the response was correct or incorrect; the frames are constructed so that at least 95% of the trainees will provide the correct answer at their first attempt; and the pace of the learning is set by the trainee. (Patrick, 1992). The second approach (Crowder, 1960) highlighted the value of allowing the trainee to make errors. This led to the design of 'branching programmes' where the trainee is able to take a number of different routes through the learning material depending on their capabilities. The characteristics of these programmes are as follows: compared to linear programmes they generally have a lot more learning material; the questions at the end of each frame are 'diagnostic', that is, the responses of the trainee determines whether they are able to move on to new information or whether more explanation is needed on the current issue - this is based on the nature of the trainee's error (Patrick, 1992). In this way branching programmes are more responsive to the individual requirements of the trainee, than linear programmes which require the trainee to progress through all the learning material in the same steps. However the success of programmed learning (or instruction) was found to be not due to the use of the two different theories but rather to the careful attention given to the development process of the training.

Goldstein (1993; p. 238 - 242) cites many different evaluation studies of these programs and Patrick (1992) highlights the basic problem with evaluating these and any other training method - what it should be compared against. The many confounding factors were listed as (Patrick, 1992):

1) Information is presented in different forms, e.g. programmed instruction is visual, whereas a lecturer uses visual and auditory modes of presentation and VEs potentially uses visual, auditory and tactile modes;

2) Motivation may be increased by the novelty of the programme;

3) It is difficult to equate fairly the amount of time spent learning in programmed instruction and other similar self-pacing technologies with the structured setting of conventional teaching courses and other similar controlled training methods;

4) It is difficult to match trainees in their ability e.g. the trainees which are learning through the lecture method will be using different learning strategies to those who are being trained using a computer based training package; and

5) More time is taken to define and structure the learning material in a programmed instruction than generally in conventional courses.

Patrick (1992) cites Hartley (1966) who reviewed 112 evaluation studies comparing programmed instruction with conventional instructions. Programmed instruction was revealed as ‘superior’ in terms of time taken to learn and test results at the end of training but re-test results showed little differences. However Patrick (1992) reports that some of the studies had small numbers, were short programmes and there was little information on long-term retention or cost benefits.

However the major contribution of programmed instruction to the training field were the basic foundations for training using computers, in particular, four of the most important features are as follows (Patrick, 1992; Goldstein, 1993):

1) *The shift from an open-loop to a closed-loop training system.* In an open-loop system the learning material is just presented to the trainee to be absorbed (e.g. the lecture method of training). In a closed-loop system the trainee is given some learning material and expected to make a response. They are then given feedback on this response and also the next part of the learning material based on their response. In this way the trainee is not overwhelmed by a huge amount of material at once and can learn the material much more quickly by not having to 're-learn' material that is already known. The computer can also store information about the trainee so that they can begin again at a later date at the stage where training was last terminated.

2) *The definition of training objectives together with a more systematic organisation of the training material.* Programmed instruction forced the designer of the training to set clear training objectives and also to consider the structure, sequence and content of the steps or 'frames' of the training programme. This systematic approach has resulted in more careful consideration of each stage of the development process and provided guidelines in which to develop future training programmes.

3) *A perspective of the trainee as an 'active' participant in training rather than a 'passive one'*. The trainee is required to be active and make responses rather than just observe the training and any interaction is preferable to no interaction (Patrick, 1992).

4) *The development of self-contained training 'packages'*. The training does not necessarily have to be delivered in a training centre or learning centre or by a trainer, programmed instruction packages were designed to be trainee-controlled so that the trainee could set their own pace of learning.

The computer revolution replaced the dedicated 'teaching machines' required for programmed instruction with computers and thus 'computer-based-training' packages were developed.

5.7.1 Computer Based Training (CBT)

There are many names and acronyms that exist for the various forms of learning which are supported by the use of standard computer technology. Computer-based training (CBT) is generally a term used to refer to computers in the context of industry and military training. Other terms which have often been used interchangeably are 'computer-based learning' (CBL), 'computer-assisted learning' (CAL) and 'computer-assisted instruction' (CAI). CBL and CAL tend to refer to educational packages and CAI tends to be used to refer to both educational and industrial contexts. The distinctions between the terms are therefore quite superficial (Patrick, 1992).

There are many different types of tasks which CBT packages are used for, some of the main areas including relevant studies that have been performed were discussed by Patrick and Stammers (1977) reviewed by Patrick (1992) and briefly presented here as follows:

1) *Perceptual identification* - A number of research studies (Swets et al, 1966; 1962; Weisz and McElroy, 1964) have provided evidence of the effective use of CBT for auditory and visual identification. A CBT system was used which generated different signals and recorded the user's response and then applied the next appropriate signal and produced a report on performance at the end. The results of the studies showed that the system provided evaluation and feedback at an optimal speed however it did not produce any better learning than simple presentation of the signals.

2) *Perceptual-motor skills* - One of the earliest CBT systems was developed in the 1960s by the Computer-based Education Research Laboratory at the University of Illinois, called PLATO (Programmed Logic for Automatic Teaching Operation).

PLATO IV was used to train the task of ‘stacking’ aircraft which involved the trainee ‘flying’ an aircraft with a hand controller while viewing simulated flight and navigation instruments on a computer monitor. Studies by Finnegan (1977), Trollip and Ortony (1977) and Trollip (1979) reported that the CBT resulted in “faster training with fewer errors in comparison with more conventional training” in transfer tests using both the GAT-2 trainer simulator and an actual aeroplane.

3) *Basic intellectual skills* - The second well-known and earliest form of CBT system is TICCIT (Time-shared INteractive Computer Controlled Information Television). TICCIT was developed in the 1970s and was aimed at concept and rule learning. However an evaluation of TICCIT for teaching basic mathematics carried out by Alderman (1978) (cited by Hartley, 1985) found that for the conventional teaching course completion rate was 50% but for TICCIT groups it dropped to 16% and the TICCIT students showed less favourable attitudes to mathematics. The interpretation was that “*programs which allow each student to proceed at his or her own pace, risk losing students unable to manage their own learning*”. However the post-test data for all the participating students showed that TICCIT groups were more than 10% better than those conventionally taught and data on problem-solving tests showed even greater benefits (Hartley 1985, p144). Other studies using the PLATO system have also shown that CBT was successful in teaching mathematics and Russian particularly in schools which were less affluent (Suppes and Morningstar, 1969). Friend and Patrick (1988) also found effective a CBT for ‘induction training’ of chemical plant trainees which involved familiarisation of appearance, location and function of different pieces of equipment. In general the expansion of the computer industry over the last twenty years, especially into the home market has resulted in numerous CBT and CBL modules widely available on a variety of subjects from astronomy, medicine and car maintenance.

4) *Management tasks* - A popular application of CBT has been in providing computerised ‘business‘ games to train managers in decision-making. The first and most influential computer-based game first appeared in 1957 published in a book

called ‘Top Management Decision Simulation’ and published by the American Management Association (D’Cruz, 1993). This resulted in many large universities and business schools researching into game technology. Now computer-based games are widely available from all the main distributors of management tools.

5) *Problem diagnosis and fault-finding* - one of the main areas for CBT has been in problem diagnosis and fault-finding especially in medical situations. This is because patients, emergency situations or hazardous equipment can be simulated so that medical trainees can diagnose the situation without the risks associated with the real situation. Secondly the trainee can be exposed to a variety of situations that they may not encounter in the real training situation but may need to know in an emergency. Research carried out by De Dombad et al, (1969) found the patient populations available in hospitals were too specialised which led them to explore alternative methods of simulation such as CBT. Thirdly the computer can provide flexibility in representing the task. For example, the task can be represented in a basic abstract form like a diagram and different layers of complexity can then be applied. Finally the system can evaluate very quickly the actions of the student and provide instant feedback which is beneficial for the trainee and the trainer.

Generally the types of tasks that are being trained by CBT are wide and varied. Studies in the area have shown some potential benefits but in most cases they do not show that CBT produces any better learning than other methods and some studies have shown that there may be potential difficulties with the ‘self-pacing’ attributes of CBT with some trainees. However there are still very few evaluation studies of CBT that exist (Patrick, 1992; Goldstein, 1993). One of the largest studies of CBT evaluation was performed by the US military almost 20 years ago (Orlansky and String (1979; cited by Patrick, 1992). The review covered 30 studies which were carried out since 1968 but varied in terms of the number of trainees, training days and subject matters. Some of the interesting points were that CBT resulted in about 30% reduction in training time in comparison with conventional training, however CBT was associated with a slightly higher rate of attrition and trainees preferred CBT but trainers did not. Therefore while some savings were identified, so were two problems - slightly higher resistance to use and in particular, the reluctance of the trainers. Goldstein (1993) recommends that studies need to be directed towards the effects of 'machine-oriented' learning environments on satisfaction, motivation and development as some research (Patten and Stermer, 1969) has suggested that adult learners do not prefer to be taught exclusively by machines.

An important evaluation of CBT is the various costs and benefits associated with the development and application. Of the evaluation studies that exist in CBT, few have considered cost benefits because it is quite difficult to perform accurately (this has already been discussed in section 5.6.9). Dean and Whitlock (1984) provide a possible list of the costs and benefits to consider, shown in Table 5.4.

Table 5.4. Costs and Benefits of Computer Based Training
(Dean and Whitlock, 1984; p. 157)

Costs	***Benefits***
Start-up	**To students**
• people	• more effective use of time
• equipment	• available when needed
	• consistent presentation
On-going	• reduced travel time and cost
• course production including	• social short-term requirements can be met
design	
authoring	• not away from place of work or home
entry into computer	• practical training made more effective
correction	• may be available any time
review	• home study may be possible
validation	
other media	**To training department**
• course presentation	

running costs of equipment	• reduced instructor time
instructor time	• easier student monitoring
time of computer personnel	• accurate student monitoring
• course updating	• less classroom space needed
	• incentive to improve courses

With regard to costs, firstly there is the personnel involved. It takes a variety of skills to produce a CBT module from analysing the training needs, designing the training, programming the computer, etc. Therefore a number of different types of experts are needed to be employed. There is the cost of the equipment, which ideally may just be the equipment already available in the organisation but this is likely to need some additional parts (e.g. specialised software, possibly more computer memory or a variety of input devices). There are also on-going costs which need to be considered for example, development time is a large cost which is dependent on a number of factors e.g. if the course is new, if there is existing material, what the content is, how long the training is to take, the characteristics of the trainees, etc. Presentation of the course will have certain overheads such as, running costs, hand-outs, cost of trainers, etc. Finally revising and up-dating the material will incur costs that must be considered.

As to benefits for the student or trainee, firstly there is more effective use of their time. For example, the self-pacing nature of a CBT allows the trainee to complete the training in a time suitable to their abilities. Trainees who are quicker at learning are therefore not hindered by the pace of the training and others who need more time for learning are able to repeat and review difficult sections in their own time. The training is also available when the trainee is ready for it as once the CBT system has been developed the trainee can have access to it at any time and as many times as they want. CBT can be available day or night and can even be taken home. This reduces the travel time and expenses that may have been incurred from going to remote training centres and does not take the trainee away from work longer than is necessary. The training material is also presented consistently so every trainee receives the same standard of training, so should be able to reach the same standards of proficiency. Special short-term requirements can be met, that is, trainees can be suddenly re-skilled by a CBT at any time and practical training can be made more effective, if for any reason it cannot be carried out in the real situation, CBT can provide a good alternative.

In terms of benefits to the training department, firstly there is more effective use of the trainer, as they may be limited in number and therefore can be used when necessary for more advanced training. Also modifications can be made more easily, as just the information that requires up-dating can be done so without having to re-design the whole training package. The computer can also monitor and record statistics on the trainees performance and provide instant reports saving a lot of time and work for the trainer. Finally the feedback from the trainee can be constantly monitored so that modifications to the training can be made when required.

5.7.2 Simulators

The very first simulators were used for flight training before and during World War I and World War II. The Link trainer developed for instruments flight training in 1929 is recognised as the first applied simulator (Rolfe and Staples, 1986; and Meister, 1995). After World War II the growth of simulator technology expanded to other areas including: automobiles, trucks, railroads, ship propulsion and collision avoidance systems, submarine and surface warfare system, air traffic control, tanks, artillery, missiles, military command control, nuclear power, mining, fire fighting and systems research (NCR, 1985; Meister, 1995).

The types of task which they provide potential advantages for, are where: tasks in the operational situation are inaccessible, there is high costs and consequences of errors; simulation is more economical than using the real situation; simulators have more support mechanisms when help is required; certain aspects of the task can be manipulated; and by being able to present different levels of complexity of the task this can reduce the stress associated with the whole task (Patrick 1992).

The design of a simulator involves all the usual steps of the training development process however the additional consideration is how to best represent the task and which features can be omitted while maintaining the ‘psychological’ fidelity of the task (discussed in section 5.4.2). Therefore the issue of ‘transfer-of-training’ is central to the development of simulators (Patrick, 1992).

Orlansky (1982) has summarised many of the studies considering the cost-effectiveness of simulators in commercial and military training. The main problem has been identifying all the elements which contribute to the cost of training in the simulator and the real world situation. In particular the assessing of the simulators ability to reduce the hazards associated with training in the real world and the simulators inability to provide training on some of the tasks which are essential in the real world but cannot be reproduced. Therefore results of cost effectiveness are very much hindered.

5.8 Discussion

This chapter provides an outline of the main features to consider in the training development process and the use of computers in training. The key stages are the training needs analysis, design and development of the method and the evaluation. Each of these stages is supported by many different theories on learning and skill acquisition which have changed over the years our understanding of how to develop effective training applications. What can be clearly seen is the difficulty in conducting any form of structured evaluation without consideration of the development process and in particular, the training needs analysis stage. This is because many of the expected outcomes are identified at this stage which can then be used to measure the effectiveness of the application. Effective training has a number of benefits to the organisation. As discussed in the analysis of organisational needs (section 5.4.1) there are a number of 'environmental' factors - general external (e.g. political, economical, social, technological); specific environmental (e.g. customers, suppliers); and internal (e.g. employees, trade unions, shareholders) - that have an influence on the state of the organisation. In order to control these influences it is necessary that the organisation develops a strategy to cope with change. One of the ways of supporting this strategy is through effective training. This is very much the belief of many organisations at present due to the Government White Paper 'Employment in the 1990s'. This expressed concerns over the low level of investment in training in the UK and the difficulties in proving the link between training and organisational success (Boydell and Leary, 1996). The *Training in Britain* survey (Training Agency, 1989) estimated that only 15% of organisations tried to evaluate the benefits of training and only 2.5% attempted any cost/benefit analysis (Bee and Bee, 1994). This led to the National Training Task Force to develop the Investors In People (IIP) initiative which is awarded to organisations that meet the national standards of training. An example of an IIP standard is shown in Table 5.4 overleaf (Boydell and Leary, 1996).

Table 5.5: IIP standards, assessment indicators and typical supporting evidence (Boydell and Leary, 1996; p.195)

Standard	Assessment Indicators	Typical Evidence
1. An ***Investor In People*** *makes a public commitment from the top to develop all employees to chieve its business objectives.*	1.1 There is a public commitment from the most senior level within the organisation to develop people. 1.2 Employees at all levels are aware of the broad aims and visions of the organisation	Mission or vision statements; written plan; letter of committment; active involvement by the organisation in relevant activities (ITOs, MCI, Business/Education partnerships); National training Award.
Every employer should have a written flexible plan which sets out business goals and	1.3 there is a written but flexible plan which sets out business goals and targets	Mission or vision statement, expressed in a way that everyone can understand; employee survey;

targets, considers how employees will contribute to achieving the plan and specifics how development needs in particular will be assessed and met.	1.4 The plan identifies broad development needs and specifies how they will be assessed and met. 1.5 The employer has considered what employees at all levels will contribute to the success of the organisation and has communicated this effectively to them 1.6 Where representative structures exist, management communicates with employee representatives a vision of where the organisation is going and the contribution employees (and their representatives) will make to its success.	employee representatives' statements; employee-briefing arrangements Relevant extracts from plan; evidence that plan has been reviewed where necessary. Relevant extracts from plan; top-level review. Mission or vision statement and what it says about people; employee survey; personal plans. Statement from employee representatives; minutes of joint meetings; material produced locally by employee representatives

The '*Training in Britain*' survey (Training Agency, 1989) also estimated that employers spent £18 billion on training in 1986/7 and sponsored 145 million training days (Bee and Bee, 1994). However a more recent survey by the Employment Department (1993) showed a decrease in training from 40 million days in 1991 to 26 million days in 1993 although this and other surveys indicated that the same number or more people are being trained (Bee and Bee, 1994). Therefore regardless of such initiatives as the IIP and the National/Scottish Vocational Qualifications (NVQs/SVQs) the amount of training per employee is decreasing.

This highlights the importance of ensuring that training applications are effective in decreasing the amount of learning time that is required by the trainee. What is apparent when looking at the development of computers and training is that computers have the potential for fulfilling this need by providing effective training as it is needed. In particular computers have meant that training is no longer constrained to training or learning centres, it can be easily distributed to any location (via networks or the mail) that has the necessary equipment. Therefore training can be accessed at convenient times and as frequently as required.

Dean and Whitlock provided a list of the potential costs and benefits provided by computers (Table 5.4) and noted that time savings can be made by more efficient use of the trainer as well as, the trainee. For example, the trainers time can be spent providing enhanced training to trainees. Given this list and subsequent comments, the advantages of using computers for training can be summarised as follows:

- Portable or can be offered through on-line network facilities
- Easily reproduced
- Provides quick and easy access to relevant information
- Allows familiarisation of product, process, place without risks to person or equipment
- Allows practice in a 'natural' way
- Provides instant and realistic consequences to actions.
- Can be re-set instantly and frequently.
- Flexible to changes in design
- Trainee can be self-paced
- Can match individual needs and requirements
- Can automatically record progress
- Can provide feedback
- Can increase interest levels and motivation of trainees to learn

These are discussed further in the next chapter. The disadvantages are mainly cost of development and the possible negative attitude some trainees may have against computers. Training by CBT is usually carried out individually therefore the trainee may feel isolated. This lack of human interaction limits the number and type of questions that can be asked and also the number of ways a question can be answered. Also there are varying abilities of computer skills. This means that some people will spend a long time learning how to use the computer rather than learning the subject matter, which negates any time savings.

As to when it is best to use a computer Seltzer (1971) suggests that when computers offer a solution not available by other methods; when it is better than an existing method but not where it exceeds potential savings.

Evaluation also has a number of difficulties as noted by Patrick (1992) mainly because of the differences in design, presentation and use. More time is taken to design a CBT because it involves a number of people - subject matter experts, designers, computer programmers, etc. - and has to include as much information as possible, to anticipate the varying abilities of the trainees and the types of questions that are likely to ask. Alternatively, a manual or delivering a lecture is likely to be by a subject matter expert or experts who are solely responsible for the content and presentation of the material. The information on a CBT can also be presented using a mix of many methods or multimedia e.g. text, video, animation, diagrams, VEs and so on. Trainees are able to spend different amounts of time on the training which may result in history effects threatening validity of the information collected. Also the trainees use different learning strategies e.g. passive in a lecture or actively taking notes.

The next chapter considers all the information gathered about the development process of training and suggests a framework in which VET applications can be developed and evaluated in a structured way.

CHAPTER 6: A framework to guide the development process of VE training (VET) applications

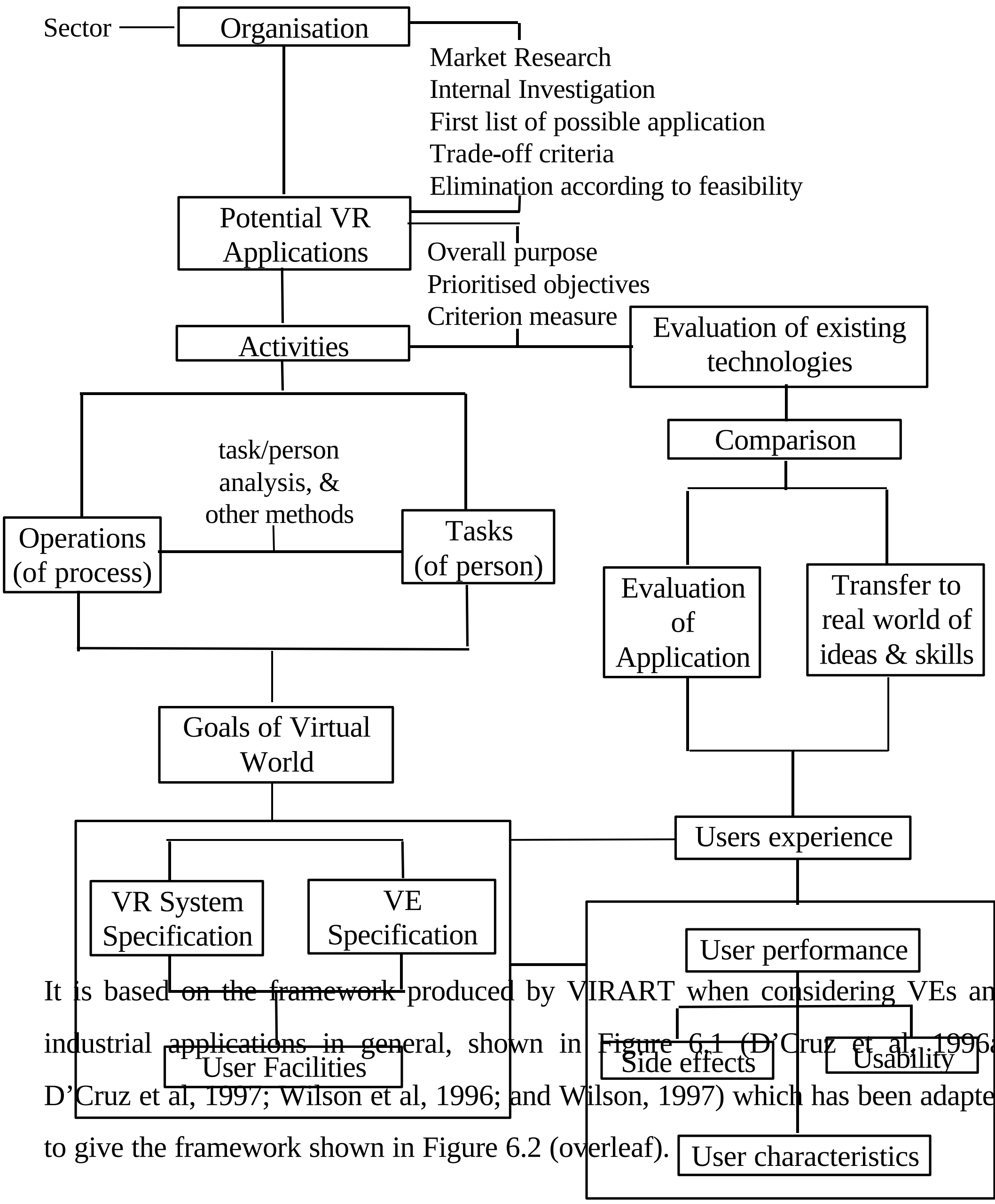
6.1 Introduction

This chapter suggests a framework in which to develop virtual environment training (VET) applications. Consideration of the evaluation process of training (discussed in chapter five) highlighted that evaluation is very much inter-linked to the training development process. Also the results of the MOVE programme (discussed in chapter three) revealed that one of industry's needs was for a structured development process in order to consider their VE applications. Therefore, an initial framework was proposed (see Figure 6.1, overleaf) which, as part of this current research, was revised in light of research into the training development process with particular focus on VET. A brief outline of how the framework was formulated is given in section 6.2 then each stage of the development process is described - the proposal stage in section 6.3, specification stage in section 6.4, building stage in section 6.5 and evaluation stage in section 6.6. The chapter ends with a discussion about this framework in section 6.7.

6.2 Formulation of Framework

The framework was formulated by the author's own experience with VEs and industry (see chapter three and four). It takes account of how industry already develop their VET (see chapter four) and the systematic development process of training programmes (see chapter five).

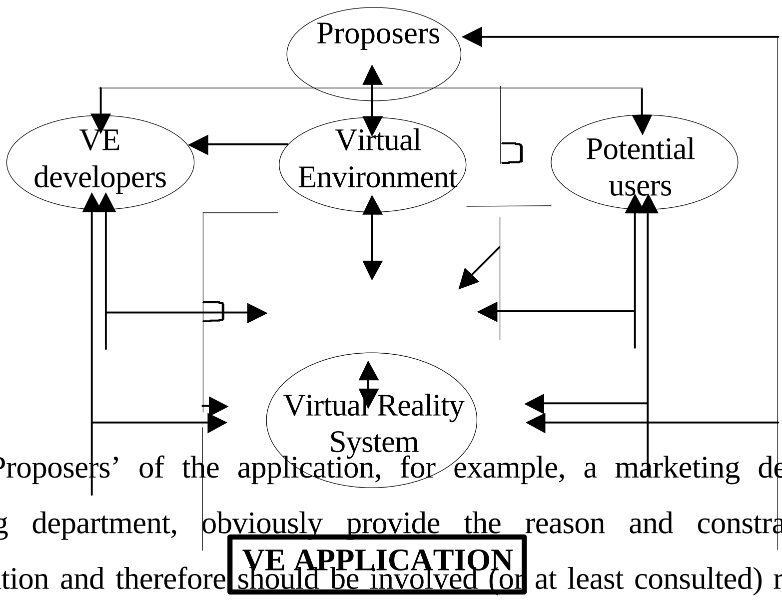
Figure 6.1: An initial framework in which to specify, develop and evaluate VE applications (Wilson et al, 1996)



It is based on the framework produced by VIRART when considering VEs and industrial applications in general, shown in Figure 6.1 (D'Cruz et al, 1996a; D'Cruz et al, 1997; Wilson et al, 1996; and Wilson, 1997) which has been adapted to give the framework shown in Figure 6.2 (overleaf).

Much of the fundamental research work in VEs has involved investigating various attributes, such as 3D visualisation (Bajura et al, 1992; Bryson, 1992; Satava, 1993), ‘presence’ and ‘immersion’ (Sheridan, 1992; Witmer and Singer 1994; Slater and Wilbur, 1995; Barfield et al, 1995), etc. Little consideration has been given to date to the overall development process for VE application. It is important to understand the influences on this process as it will effect the quality and usefulness of the eventual application. For example, different VE developers are very likely to produce different VEs from the same specification; this has been seen in VIRART’s own laboratories. This is because VE developers apply their own individual experiences, capabilities, styles and imagination to translating the specification into the appropriate VE, in much the same way any designer does. So decisions on factors such as layout of VE; numbers of objects; complexity and detail of objects; interactive elements and so on, may vary. There are five broad components which will influence the outcome of a VE application as shown in Figure 6.3 below.

Figure 6.3: The components that influence a VE application



The ‘Proposers’ of the application, for example, a marketing department or training department, obviously provide the reason and constraints of the application and therefore should be involved (or at least consulted) right through the process. In relation to the other components, it is the proposers that specify who the VE developers and potential users are and in agreement with the developers, decide what should or should not be included in the VE and what VR system will be used.

The nature of the VE developers has a direct relationship with what the design of the VE will be like. As mentioned before, each VE developer influences design with their own experience, capabilities, style and imagination. To a certain extent their individual characteristics, in particular their knowledge, skills, abilities and attitudes will influence the type of VR system chosen but resources of the proposers will have a major influence also.

The potential users will also have characteristics which will effect the design of the VE and the choice of VR system. For example, if the VE application was to be used to train engineers and non-engineers in a number of maintenance tasks, then the design of the objects in the VE for the engineers need not include precise detail, as it can be assumed that they will be able to easily distinguish the equipment. However for the non-engineers it is possible that greater detail is required so that they will be able to recognise the equipment when transferring to the real world situation. Also the potential users influence the configuration of the VR system used. If most of the potential users are likely to have mobility problems for instance, then a desktop VR system with adapted input devices may be the most appropriate.

Finally the planned design of the VE should influence the choice of VR system; for example, if high quality photo-realistic objects are required then a high-end system like the Silicon Graphics' 'RealityEngine2' may be necessary (although in most cases the proposers choose the system first and then the VE is designed).

Influences of all of these components on the VE application must be considered. In particular if the VE application is not successful, it could be the result of problems within one or *all* of the above. For example, the proposers may have inadequately specified what they actually want, the VE developers may have wrongly interpreted this specification and designed the VE inappropriately, the VR system chosen may be hindering the users in fulfilling their objectives and the profile of the potential users may have been misleading. Therefore it is a good idea to consider carefully each stage of the development process to avoid misunderstandings.

6.3 Framework for the development of VET applications

As shown in Figure 6.2, the framework for developing VET applications has the main stages of the training development process (discussed in chapter five). The process begins with some training needs analysis which involves analysis of the performance problem at three levels - organisational, job and person (Goldstein, 1993). The information from this analysis then provides much of the detail for the next part of the process - the design and development of the training. This also involves three components - the training content, the trainee and the training methods and strategies (Patrick, 1992). The training content and the characteristics and requirements of the trainee are usually quite clearly defined by the training needs analysis stage, however the structure of the training and the delivery method are still open to interpretation. At this point of the training development process, VET can be proposed as a possible choice of training method. Therefore the development process of VET will now be discussed.

6.3.1 Proposal stage

There are many different training delivery methods to choose from - instruction manuals, training videos, simulators etc. - all of which have their various advantages and disadvantages. These are usually well known and documented (Goldstein, 1993). However, VET is still a new technology and applications are not, at present, available ‘off-the-shelf’. In fact, if a trainer decides to use VET then it is possible that a lot of investment in time and money will be required to develop an appropriate application. Therefore the first question which needs to be considered is whether or not VEs are potentially valuable for providing a solution to any particular training need.

6.3.1.1 Is VET potentially valuable?

In order to make this decision the problem or task must be considered against some form of criteria. In order to develop these criteria, the attributes of VEs (as discussed in chapter two) need to be explored, in terms of how they match some of the training needs of industry for computer technology. These needs have already been expressed to some extent by those companies who are already looking at VET (discussed in chapter four) and also highlighted in previous computer based training (CBT) applications (discussed in chapter five). This is shown in Figure 6.4 (overleaf).

The attributes of VEs mean they can potentially offer the same advantages as other computer-based technologies. They can be portable or provided through on-line networked facilities and are easily reproduced. This becomes essential if there is a need for training a large number of people and/or they are distributed over a large geographical area. They can also make finding relevant information easier and quicker which is important if people need to frequently refer back to the information. For example, in the case of maintenance training where the trainees may not have to use their trained skills for a long time and so when they eventually need them a VE or other computer delivery method could be an effective memory aid.

Other advantages of VET are particular to VR technology. The ability of VEs to represent real and abstract three-dimensional environments suggests that people can familiarise themselves with a product, process or place without risking themselves, other people or any of the equipment. Also as VR systems can support various devices that engage different senses (visual, auditory, tactile) the trainee can essentially experience the learning environment in a similar way to the working environment. This is essential if the working environment is particularly dangerous or is situated in a remote or hazardous place. Also the real equipment may be too expensive to reproduce or to be taken out of the working process or to let novices handle it. It may be too large or too delicate to move and therefore needs to be represented in some way.

Figure 6.4 : Matching the attributes of VEs with the training needs of industry for computer technology

Attributes of VEs

- *computer-generated representations of real or imaginary environments;*
- *experienced as three dimensional via a number of senses - visual, aural and/or tactile;*
- *objects within these environments are independent of the user and can display real world behaviour;*
- *the user or users have autonomous control - the freedom to navigate and interact with objects, using a number of different viewpoints;*
- *interaction occurs in real-time; and*
- *the users experience feelings of presence and/or immersion.*

Training needs

- VE training programme is portable, can offer on-line networked training facilities, is flexible to changes
- allows familiarisation of a product without risks to person or equipment
- allows user to take control of learning pace, to monitor own progress by receiving continual feedback, to carry out procedures as realistically as possible and in real time, therefore procedures can be timed.
- can engage a number of senses (including movement), can meet individual needs, can automatically record progress
- allows familiarisation of a situation that maybe too hazardous, too remote, not even developed yet, etc.
- it is a new technology, engages interest thus increasing motivation

The capability of objects in the VE to behave as they would do in the real world and for the trainee to activate them as you would in the real world, allows the trainee to 'virtually' practice the trained skill and receive instant and correct feedback on the consequences of their actions. This serves to reinforce the training the same way that practising on the real equipment does but without the anxiety of potential injury or damage. The VE can also be re-set to the initial state of the equipment instantly and as many times as required.

VEs are also flexible to allow small and large modifications, according to changing needs and changing design. This becomes necessary especially if the equipment or system for which training is planned has not yet been built and is still in its design stages. Moreover the VE is not constrained to a pre-defined path; so as long as the objects have sufficient complexity the trainee can try out a number of different methods to produce the same outcome. This allows the trainees to find the method that best suits them. Alternatively, the VE could be programmed to be sensitive to their individual needs. The trainees have control over their pace of learning and the VE could be designed to match their abilities. This progress can be tracked and an automatic record kept for future reference. Also automatic feedback can be provided, for example, a scoring sheet recording the number of successful actions and number of errors.   This allows the trainee to monitor their own progress, as well as allowing the training manager to make decisions about the training programme and the trainee.

Finally, VEs offer the training manager a choice of a new innovative delivery method. If nothing else its value may just lie in the fact that it increases the motivation levels of the trainees. An enjoyable training experience suggests that trainees will have a more positive attitude to the learned task and will be encouraged to use their training on-the-job.

Therefore the attributes of VEs potentially have much to offer training applications and it is possible, from matching these attributes to suggest some form of VE selection criteria (D'Cruz et al, 1997). At the time of writing this work, others have also suggested similar selection criteria and two of these are discussed next.

6.3.1.2 Media Selection Model

Dorrington and Elliot-Square (1995) of Westland System Assessment Limited (a consultancy group in training, operational studies and logistics mainly for the defence industry) have developed a 'Media Selection Model', including four VR options. These are 'Desktop Virtual Reality - standalone or networked', and 'Immersive Virtual Reality - standalone or networked'. The model is based on the Automated Instructional Media Selection (AIMS) model developed by the U.S. Navy (cited by Dorrington and Elliot-Square, 1995). It consists of a table with the vertical axis listing seventeen types of training media and the horizontal axis listing different types of training objectives. These objectives are divided into seven categories - type of learning, display characteristics, response mode, evaluation mode, environmental conditions, learning level and special requirements. At the corresponding cell for between a training objective and a type of media, a number between '0' and '5' has been given. This indicates on a scale, the relevance of the media to the training objective; '0' means the media is definitely inappropriate and '5' means that it is highly appropriate. This model is shown in Table 6.1.

Table 6.1: Media Selection Model (Dorrington and Elliot-Square, 1995)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
Types of Learning																	
Cognitive	2	4	5	5	5	5	5	2	3	3	4	0	1	1	1	1	1
Psychomotor	0	0	0	3	3	3	2	0	2	3	3	5	5	5	5	5	5
Affective (Attitudinal)	1	4	3	3	4	4	5	0	0	0	2	0	0	2	2	2	0
Display Characteristics																	
Verbal	0	5	4	3	3	3	4	0	0	0	0	0	0	0	0	0	0
Text	5	3	3	4	4	4	4	0	0	0	0	0	0	0	0	0	0
Photographs	3	5	4	2	4	4	4	0	0	0	0	0	0	0	0	0	0
Drawings	5	5	4	2	4	4	4	0	0	0	0	1	0	0	0	0	0
Diagrams	5	5	4	2	4	4	4	3	3	3	3	1	0	0	0	0	0

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
Colours	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Visual Motion (constant)	0	0	4	2	4	4	4	4	4	4	4	0	2	3	5	5	5
Visual Motion (variable)	0	0	0	2	4	4	5	5	5	5	5	0	2	3	5	5	5
3 Dimensional	0	0	0	3	4	4	5	5	5	5	5	5	5	5	5	5	5
Audio (Voice)	0	0	4	0	5	5	5	3	4	3	4	0	3	4	5	5	4
Audio (Ambient)	0	4	3	0	3	3	5	5	5	5	5	0	2	3	5	5	5
Tactile Cues	0	0	0	0	0	5	0	0	0	0	0	0	5	5	5	5	5
Kinesthetic Cues	0	0	0	0	0	0	1	1	1	1	1	5	5	4	5	4	5
Response Mode																	
Verbal/Written Response	5	3	0	0	2	2	2	0	1	0	1	0	0	0	0	0	0
Voice Response	0	5	3	0	3	3	2	2	3	2	3	4	4	4	4	4	4
Decision Indicator	0	4	0	5	5	5	5	5	5	5	5	1	3	3	3	1	0
Fine Motor Response	0	0	0	0	0	5	0	0	0	0	0	5	5	5	5	5	5
Gross Motor Response	0	0	0	0	0	1	0	0	0	3	3	5	2	3	5	3	5
Tracking Response	0	0	0	4	4	4	4	4	4	4	4	1	4	5	5	3	5
Manipulative Response	0	0	0	0	2	4	2	2	2	3	3	3	3	5	5	5	5
Continuous Response	0	0	0	0	0	1	1	1	1	0	0	2	4	5	5	2	5
Evaluation Mode																	
Instructor Evaluation	5	5	3	1	1	1	1	1	1	2	2	4	4	4	4	4	5
Self Evaluation	5	0	0	0	0	0	4	4	4	4	4	4	4	4	4	4	4
Peer Evaluation	2	2	2	0	0	0	0	0	0	3	3	1	1	3	4	4	4
Automated Evaluation	0	0	0	4	5	5	5	3	3	3	3	0	3	4	5	3	3
Verbal/Written Evaluation	5	5	0	0	2	2	5	0	0	0	0	1	0	0	0	0	0
Immediate on Response	2	3	0	3	5	5	5	3	3	3	3	1	3	4	4	2	2
Immediate on Error	1	0	0	3	5	5	5	3	3	3	3	1	3	4	5	2	2
Post Session Evaluation	1	5	0	5	5	5	5	5	5	5	5	1	3	4	5	1	2

where letters correspond to:

a) Workbook
b) Instructor with Vugraphs/Drawings
c) Electronic Audio/Visual Training
d) Interactive CBT
e) Interactive CBT with Multimedia
f) Interactive CBT with Emulated HCI
g) Mutimedia, CBT and Desktop VR
h) Desktop VR: standalone
j) Immersive VR: standalone
k) Immersive VR: networked
l) 3D Model
m) Psycho-Motor Trainer: Procedural Level
n) Psycho-Motor Trainer: System Level
o) Psycho-Motor Trainer: Mission Level
p) Real Equipment: Disembodied
q) Real Equipment: In-Situ

i) Desktop VR: networked

Table 6.1 (continued): Media Selection Model (Dorrington and Elliot-Square, 1995)

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
Environmental Conditions																	
Physical Motion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	5
Acceleration	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	5
Pitch/Roll	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	5
Vibration	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	5
Restrictive Clothing	0	0	0	0	0	0	0	0	0	2	2	0	4	4	4	4	5
Learning Level																	
Familiarity	5	5	5	5	5	5	5	5	5	5	5	5	1	0	0	1	0
Perform Procedure	0	0	0	2	2	4	2	2	2	3	3	5	5	5	5	5	5
Perform Job	0	0	0	0	0	3	3	3	3	4	4	1	3	5	5	3	5
Perform Mission	0	0	0	0	0	2	3	2	2	4	4	0	1	3	5	1	5
Special Requirements																	
Crew/Team interaction	0	0	0	0	0	0	0	0	4	0	4	3	0	2	4	0	5
Large Memorisation component	3	3	3	5	5	5	5	5	5	5	5	0	2	2	2	2	3
Random Access Logic	1	3	0	5	5	5	5	5	5	5	5	1	2	2	2	2	3

where letters correspond to:

j) Workbook
k) Instructor with Vugraphs/Drawings
l) Electronic Audio/Visual Training
m) Interactive CBT
n) Interactive CBT with Multimedia
o) Interactive CBT with Emulated HCI
p) Mutimedia, CBT and Desktop VR
q) Desktop VR: standalone
r) Desktop VR: networked
r) Immersive VR: standalone
s) Immersive VR: networked
t) 3D Model
u) Psycho-Motor Trainer: Procedural Level
v) Psycho-Motor Trainer: System Level
w) Psycho-Motor Trainer: Mission Level
x) Real Equipment: Disembodied
y) Real Equipment: In-Situ

So once the training objectives applicable to the training task have been identified, a mean score between 0 - 5 for each type of media can be calculated. This potentially identifies the most appropriate choice of technology.

Many of these potential benefits have already been discussed in the previous section but not the possible limitations. Firstly Dorrington and Elliot-Square (1995) say that VEs are not necessarily appropriate for the 'Affective or attitudinal' type of learning, that is, changing attitudes and ideas. However this appears to be contradicted by the number of applications in the therapy area on phobias and rehabilitation. VEs are being used to help people re-learn basic living skills e.g. making a cup of tea, getting dressed, and so on (Brown, 1996; Brown et al, 1997) and overcome their aversion to spiders and pain (Carlin et al, 1997; Hoffman, 1997; and Hoffman et al, 1998).

As for display characteristics, if colours are important then VEs can use realistic colours, as well as textures that can make the appearance of objects more realistic. These colours can also respond to changes in light and viewpoints to produce the appropriate effects and shadows. As for visual motion, objects can move in VEs in potentially complex and realistic ways independent of the user and can be experienced as three dimensional. Dorrington and Elliot-Square (1995) however, say that VEs are inappropriate for displaying verbal information, textual information or showing photographs or drawings. This is not to say that VEs cannot show all of this information, but rather why use this technology if real text, photographs and drawings are sufficient for learning. An area which is currently limited in VEs is tactile and kinaesthetic cues, that is, providing the sense of touching and feeling objects. Work is still on-going to provide these cues (see chapter two).

When considering the most appropriate response mode, VEs can be used as a 'decision indicator' that is, they can respond to a decision with the relevant consequences. They can also provide a 'tracking response'; that is, monitor the trainees' reactions to situations that are constantly changing or are unpredictable e.g. driving along a road. However Dorrington and Elliot-Square (1995) also say that VEs are inadequate at providing verbal or written responses and again this is not strictly true. The VE can be programmed to provide such responses, and although these are not as flexible as the responses provided by an instructor or as comprehensive as a manual, but they are still possible. As for fine motor response this is difficult in current systems because of the problems with tactile and kinaesthetic cues. Also continuous response can be hindered by the speed of the computer.

In terms of the modes of evaluation, as discussed before VEs can provide self-evaluation in the forms of instant feedback as trainees perform tasks in the VE and at the end of a session where a type of score sheet can give them some indication of how well they did. VEs also offer good post-session evaluation as records are automatically kept and the trainer can review performance and the subsequent outcomes at any time. The only type of evaluation where Dorrington and Elliot-Square (1995) say that VEs may not be as appropriate is in the verbal and written form. However, again this is not to say that VEs cannot provide this type of evaluation but rather that to engage a VE to do so can be very costly and time-consuming and a waste of the resource when other methods would be more appropriate.

When considering environmental conditions, VEs can allow trainees to familiarise themselves with a product, process and place. However such conditions as physical motion, acceleration, pitch/roll and vibration are very difficult to provide for the trainee to experience with current VR systems.

Finally, for special requirements, if the trainee has to memorise a lot of basic facts then VEs allow them to do so through self-paced learning which is repeatable.

Also VEs have a certain amount of 'random access' so trainees are able to explore a number of different actions and see their outcomes.

Dorrington and Elliot-Square's (1995) model highlights some of the benefits of VEs compared to other methods of training and also some of the limitations, but a few of these are misleading. The ratings are subjective but based on experience in advising and developing VE applications up until 1995. Therefore it is not surprising that some of this information is unclear when considering the specification of the computer systems available in 1995 and the types of applications being envisaged for VEs. Possibly the model needs revising in light of current systems and applications - this is the dilemma of basing assumptions on technology, especially computer technology in the current climate. Also testing of the appropriateness of this scoring system is still needed.

6.3.1.3 Virtual Reality Training Decision Tool

Another model with particular relevance to this research, as it is focused specifically on VET, was produced by Adams Consulting Group, Inc. and the Research Triangle Institute (RTI) both from the U.S. and released in 1997 over the World Wide Web. This also provides a method for the trainer to 'quantify' the decision to use VEs. The model consists of a questionnaire with four sections: Learner profile, Learning content, Learning Objectives and Computing Resources. Under each of these headings is a number of questions with multiple answers that have corresponding numbers of points (0, 5 or 10 points). The answer selected determines the number of points gained. When the questionnaire is completed the points are added up and the 'recommended action' is given. For scores over 200 points it is suggested that "VR Multimedia Training can be very beneficial to increase the effectiveness of your organisation's training". Adams (1997) uses the term VR Multimedia to refer to a desktop VR system supporting a VE with multimedia accessories e.g. text, pictures and video clips which can be activated within the VE. For scores between 150 to 199 points it is suggested that "*VR Multimedia Training should be investigated as an effective solution. Review the* *sections of this tool for potential barriers and opportunities*". Finally for scores between 0 to 149 points it is suggested that "*VR Multimedia Training may not be the best option for your training needs at this time. Use the VR Multimedia Decision Tool in the future to see if conditions change*". (A copy of the complete questionnaire can be found in Appendix II.) From this questionnaire, VEs score full points in the following cases, shown in Table 6.2 (overleaf).

Again many of these issues have been discussed before, except that this model considers whether the trainee has to use a computer in their job and whether the company already has access to computer resources. This is important, as using a computer for training may ease the transition for those people who are uneasy about using computers and for those who are already 'computer literate' the technology will be more easily accepted. Also by already having some of the hardware it minimises the cost of using VEs.

Table 6.2: High scoring areas for VR systems from the Virtual Reality Training Decision Tool (Adams Consulting and RTI, 1997)

SECTIONS	HIGH SCORES
Learner Profile	• there are over 50 trainees • more than five training sites • the distance the trainees have to travel requires more than one over night stay • it is preferable to let trainees set schedules • the trainees are exposed to unacceptable risks when they demonstrate critical skills • preferred learning style is for independence • the trainees are required to know how to use a computer as part of the job or anyway • trainees have widely different skill levels
Learning Content	• content already available in interactive form must be developed to meet requirements • type of learning content requires hard skills e.g. technical • learning is most effective with interaction with the physical environment • most effective learning of interaction with physical objects or equipment requires manipulation of the objects or equipment • If actual physical environments or objects are not available to the trainee • if the training content includes topics which are potentially hazardous • if content questions must be answered and experts can be made available • if content contains concepts which are difficult to grasp
Learning Objectives	• If performing skills or tasks rather than just learning about them is essential

	• If the ability to practice skills and rehearse procedures is necessary • if consistency is of content covered is very important • if tracking and evaluation of skills to be performed is very important • if performance training across multiple courses is required • if the program will be offered more than 19 times • if frequency of integrated updates is more than 6 months • if development time of training can be than 6 months
Computing Resources	• If computer hardware at learner site is available • if hardware and operating system at learner site is identical

6.3.1.4 Virtual Environment Selection Criteria

Considering the attributes of VEs outlined by this researcher and shown in Figure 6.3 and the two models presented above some form of VE selection criteria can be suggested. As there is as yet, little evidence to support a quantified approach, these are in the form of a checklist to provide the trainer with some idea of the capabilities of VEs. These proposed criteria are shown in Table 6.3.

Table 6.3 Virtual Environment Selection Criteria

CRITERIA FOR SELECTION	"3" if yes
The Potential Trainees	
• there is a large number of trainees to be trained	
• the trainees are distributed over a wide geographical area	
• the trainees are not available at the same times	
• the trainees need frequent training	
• the trainees have different knowledge and skills	
The Task To Be Trained	
• 3D visualisation is important for learning the task	
• hands-on interaction is important for learning the task	
• freedom to explore the product, process or place is an effective way of learning	

the task	
• self-paced learning is an effective way of learning the task	
• current methods of training are inadequate	
• there are potential cost savings with VET compared to other methods of training	
• There is a limited number of instructors/experts	
The Working Environment	
• the product, process or place is too dangerous for training	
• the product, process or place is still under development and unavailable for training	
• the product, process or place is too expensive to be used for training	
• there is limited access to the real product, process or place for other reasons	

This VE selection criteria considers the potential trainees, the tasks to be trained and the working environment. Under each of these headings are the areas where VET can make a difference. The trainer must first consider each task that needs to be trained and apply the checklist, as for some tasks within an entire job other methods of training may be more appropriate. This was discussed in chapter five, where Patrick (1992) suggested that where a problem in performance is identified, there are four possible solutions - training; personnel selection; ergonomics (job re-design) or a combination of all three. Then, having examined the importance of these areas an informed decision on whether to proceed with VET can be made.

This checklist considers the attributes of VEs disregarding the type of VR system used, as system technology is constantly changing. Therefore it can be applied regardless of the current state of VR and until more is discovered about VEs. It is meant to be a useful starting point in deciding whether VEs have potential value for a particular task. It is not a quantified approach in the way that the more boxes which are chosen, the more likely VEs will be of use as each feature has a different priority attached to it. For example, even if only one category is selected but this is a fundamental issue like "it is too dangerous or hazardous to use the real equipment or process for training the task", then VEs could be a potential solution and therefore should be considered along with other options.

Therefore the issue of when it is appropriate to use VEs is still an area that very much needs addressing especially as there is limited research and few examples of working applications to support potential benefits.

6.4 Specifying VET applications

The specification stage involves making decisions about the application, this relates to the 'Training Specification' discussed in chapter five, which is produced after the training needs analysis stage of the development process. In developing VEs this training specification becomes extremely important as it provides most of the information required by the VE developer about the application. The parts of this specification have already been outlined in chapter five therefore it is discussed here with particular reference to the information it provides to the VET process.

(i) Background to the organisation

Background to the organisation should outline the main 'Proposers' of the application so that the VE developer is aware of who makes the decisions.

(ii) Description of the target training population

Knowledge of the target population or 'end-users', in particular their existing knowledge, skills, abilities and attitudes will aid the VE developer in knowing at what level of complexity to design the VE and what other features may be required. For example. if the trainees were non-experts then they would require quite detailed views of the working environment and more help or support facilities, then trainees who have some level of skill already. The VE developer also needs to know how much access they have to end-users in order to test various stages of the design.

The overall aim of the training is linked to the organisational need (identified at the

inning of the process). Knowledge of the

organisational need provides the driving force behind the application and gives the

VET outcomes. For example, to 'increase productivity' will be of interest to the

interest to the personnel department, etc.

The objectives should give an indication of how the aims will be fulfi

worth considering these in terms of what tasks are to be performed, in what conditions and to what acceptable standard of performance (Mager, 1991).

measure the VET application against at the end of development. It is worth noting

likely to change as more is understood and learnt. It is important that the proposers of the VE applicatio

should be constantly reviewed in light of the capabilities and limitations of the VR system, the VE and the VE developer.

(v) Choice of VR system

f the application (from

the training needs analysis stage) and the proposed design of the VE. In most

computer resources already available in the proposer's organisation. As discussed

- desktop VR,

these vary depending on the number of peripherals required i.e. visual displays, input device area, choice tends to be either desktop VR or headset VR because these systems are better known and more commercially available.

Headset systems are said to provide a greater sense of immersion as they physically block out external distractions and therefore promote a greater sense of presence within a VE. Also if the trainee needs to be aware of their perspective against the 'virtual' equipment then a headset is appropriate as it requires you to physically look up and around the VE. However these systems have given rise to considerable research work in the area of side effects associated with wearing a headset (Wilson, 1996; Nichols et al, 1997; Wilson, 1997; Cobb et al, 1998) and therefore the proposers or their advisors must be aware of this when deciding on this system. The most widely used VR system is desktop VR because it requires very little initial investment (assuming the company already uses computers) and it has fewer usability issues associated with it. Although it can be limited in its capabilities compared to other VR systems. The most reliable way of making the appropriate decision is to be certain of exactly what the training requirements are and then to see demonstrations of as many VR systems as possible.

(vi) Choice of VE Developer

The choice of VE developer considers whether the application is being developed 'in-house' or externally. Adams Consulting Group Inc. and the Research Triangle Institute again provide some recommendations for this. They suggest that high scores in the following areas imply that the company is ready to implement VEs in-house:

- If the staff assigned to the project have implemented a project using object oriented design
- If the staff assigned to the project have designed and implemented a 3-D project
- If the staff assigned to the project have used object oriented programming tools
- If computer hardware to produce materials is available
- If staff has successfully managed a technology based project

- If existing trainers can be used on VR projects
- If time available for design and development is more than 12 months.

In order to decide it is important to consider the available resources - money, time, people, equipment. In-house development should only be a feasible initial option if: the budget and time are not necessarily limited; there are people within the company that already possess some computer skills; these people are able to be trained and will be given the time to learn the system; there is equipment that is available to be used; and the company is likely to require a lot of VE applications in the future. If none of these conditions are available then it is probably best to consider advising an external group. No matter who the VE developers are, the proposers of the application should be actively involved throughout the entire development process so that a satisfactory outcome is achieved in as short a time as possible.

(vii) How the training is to be evaluated

It is necessary to consider how the training will be evaluated early on in the project in order to provide the VE developer some measures to consider the VET application against throughout the process. These should come from the training objectives discussed above (section iv).

(viii) Time-scale of the delivery of the training

The time scale is in terms of development time and the amount of time the training is likely to have within the overall training programme. For obvious reasons the VE developer must have some indication of when the final version of the application has to be completed. Although generally the time taken to develop a VET application always appears to take longer then first perceived. Some developers have warned against 'scope-creep' (I/S Analyzer, 1997). This is where the proposers of the application keep extending their requirements of the application without taking into consideration the implications on time.

(ix) The learning environment

The learning environment will have implications on the type of system chosen. If the organisation have only a small amount of space reserved for training then it limits the size of the system. For example, desktop VR requires room for a desk and a couple of chairs but headset VR requires a room that is separate from other computer systems to minimise interference with the trackers.

(x) Any other constraints

For VEs it is necessary to consider priorities of the VE. This is because of the limitations of the current technology (chapter two). Current computer systems have a certain amount of processing power and even though the boundaries of this power are constantly being pushed further and further, it is still not enough to create fully functional and visually realistic environments while maintaining a good quality of movement. For example, consider an ordinary object like a 'Bic' pen on a table. In terms of visual programming, the outer casing and the pen top must look like reflective plastic and reflect the appropriate light, shadows and objects. The outer casing is a hexagon which is not difficult to create, but it is clear and the pen inside must look like a separate part. The pen top is smooth and curved which can be difficult to create as most software uses polygons (i.e. a closed plane figure consisting of three or more straight sides that connect three or more points with no sides intersecting) to build up objects, making curved surfaces complicated to generate. In terms of object behaviour, the pen is made of five parts - the pen top, the outer casing, the bottom stop, the nib and the ink casing - all of which can be detached from each other. Also the 'pen' needs to know that unless the pen top is removed the ink cannot flow and only when the pen is held at the appropriate angle can the user write with it, then only horizontally in two directions not vertically and only when force has been applied, and so on. From this you can see that even the simplest of objects can become very complicated to model and therefore the VE developer must consider carefully what is relevant to fulfil the overall aims and objectives of the VE and prioritise efforts. These priorities will come from the aims and objectives discussed above (sections iii and iv).

As for other constraints, the most obvious influence is 'money'. Money will have a major influence on the whole project, as well as, the rest of the resources available for the project. There is no doubt that the cost of VEs is presently high in terms of money and risk. Interestingly it is not the equipment that costs the most, as is usual for other types of training methods, but it is the money required for the development time and the people for an uncertain outcome. Therefore to have some control over the costs of developing a VET application reasonable constraints must be placed on the project.

(xi) Content of the virtual environment

Further to the basic design of a training specification (defined in chapter five) the VE developer requires a description of at least the primary requirements of the VE, including:

- the objects to be included i.e. tools and equipment relevant to the purpose of the application;
- the functionality to be included i.e. what object 'behaviours' or processes must be modelled;
- the user interactions i.e. what will the user be required to interact with in order to fulfil the purpose of the application; and
- any special requirements e.g. are there alternative viewpoints that would provide better understanding of the processes?

This information is likely to change during the review sessions with the VE developer and proposer. In most cases, our evidence at VIRART is that the expectations of the proposer far exceed the actual capabilities of current VR systems, so some compromise has to be made. The key is the overall aim and objectives of the application.

6.5 Building VET applications

The building stage involves translating the specification into an effective design of VE to fulfil the aims and objectives of the application. There are currently no standard methods of designing effective VEs although work is on-going in this area (Eastgate et al, 1997). Generally VE developers have there own methods based on their knowledge, skills, abilities and attitudes and dependent on the capabilities of the VR software. However there are a number of similar stages - information gathering, development of a design or 'storyboard', modelling of the VE and then continually review and modification sessions until the application has reached an agreed level of development. These are discussed briefly as follows.

6.5.1 Information gathering

This involves the VE developer and instruction designer gathering all the relevant information from the specification and the working environment required for the VE. This information includes: dimensions, components, operations, etc.; from the real environment and supporting documentation.

6.5.2 Development of a design

The next stage involves considering the design of the VE and the sequence of events. This is referred to as 'story-boarding' by developers (Eastgate et al, 1997) and it involves defining various scenarios in picture form like a storyboard illustrating the sequence of various events to be included in the VE. It is based on the information provided by the task analysis (see chapter five) performed during the training needs analysis stage. This gives an indication of the focal points (the main objects in the environment) and other background features that may be required. When the proposer has agreed on the design and scenarios for the VE, the VE developer then models the VE.

6.5.3 Modelling of the VE

Before modelling the VE, the VE developer must make decisions on issues like (Eastgate et al, 1996):

- the visual detail of the objects e.g. size, shape, colour, texture;
- the physical properties of the objects e.g. mass, gravity, friction;
- the relationship between the objects e.g. collisions, boundaries, links, priorities; and
- any sequence of events or scenarios e.g. the lamp only works if it is plugged in and switched on, etc.

Also what background features should be considered in terms of the layout of the VE, the overall appearance (e.g. light, shadows, textures etc.), and the position and orientation of the objects. A certain number of background objects should be included in a VE, not only to make it interesting and inviting but also to aid navigation and space recognition.

Finally the user's interaction with the VE needs very careful consideration (Eastgate et al, 1996;1997). This includes issues like:

- how best to represent the user e.g. a person, an object, etc.;
- what viewpoints to include e.g. egocentric/exocentric, worm's eye/bird's eye;
- what degree of interaction with each object is necessary/possible e.g. some objects could be 'moved', altered in some way, 'turned on' or activated;
- what cues for interaction are required e.g. visual - flashing arrows to indicate where to start up the machinery; auditory - a warning noise to tell the user which areas should not be handled; textual - a text box to help the user remember a piece of information.

6.5.4 Review and modification sessions

Finally the success of the VE relies on continuous feedback from the relevant personnel - the proposers, subject matter experts and potential users - in balancing the requirements of the application with the current capabilities and limitations of the VR systems. In most cases the requirements will exceed the capabilities, and adjustments to the initial aims and objectives will have to be made.

6.6 Evaluating VET applications

Evaluation in various forms should take place throughout the development process of VET, so that the design is constantly being checked against the aims and objectives, and revisions are made before the application is implemented. As discussed in chapter 5, evaluation is divided into *formative* and *summative* approaches. Formative evaluation examines 'process' information i.e. information to identify revisions needed in the design of the application; whereas summative evaluation explores 'outcome' information i.e. information to assess the success of the application for its intended purpose. The methods involved in these types of evaluation have already been described in chapter five and are discussed below with particular reference to VET.

6.6.1 Formative Evaluation

As there are no standard ways of building VEs and therefore no real measures to assess them against, this type of evaluation is extremely important. This is because the methods specifically look at the design of the 'instruction' and identify modifications which are required to make it 'effective, efficient, useable and appealing' (Tessmer, 1993).

6.6.1.1 Self-evaluation by the VE developers

This is usually the initial approach to evaluating an application and is carried out frequently when the application is in its 'rough' stages. This requires the VE developers to 'emotionally' detach themselves from the application and examine it against the specification, for obvious errors. This approach begins as early as the storyboarding stage of the building process and the VE developers will tend to automatically self-evaluate the application right through to implementation stage. However it is advisable that the VE developers keep some form of 'development log book'. This could just be a notepad where they can write the answers to the above questions and general comments. The advantages of this are that firstly the VE developers will have a personal record of their efforts in case they come across a similar problem again (or in case they have to justify their efforts to the proposers!). Secondly, it makes review sessions with the proposers more productive, as the VE developer will have a list of specific matters that they need to discuss.

6.6.1.2 Review Sessions with Proposers

The proposers of the application must work closely with the VE developers all through the development process to ensure that their requirements are being met. The proposers must be aware of the capabilities and limitations of the chosen VR system and the implications these have for their requirements. In this way early decisions, rather than too late, can be made on whether to change the design of the VE, the VR system being used or even the requirements themselves. Review sessions with the proposers of the application should be carried out frequently and can be made more productive if structured in the following way. Firstly the VE developer should prepare specific questions from the outcome of their own self-evaluation of the VE (see section before). In particular for areas that they are uncertain about or areas that they need confirmation on. Then they should demonstrate what has been modelled so far and record the initial feedback information from the proposers. If there are any questions that have not been answered from the VE developer's own specific list, these should then be asked. The review should finish with some agreement of the actions to be taken before the next review meeting.

6.6.1.3 Expert and Non-Expert Reviews

These should be carried out once the VE developers and the proposers have decided that most of the application is complete but the quality of the information and system need to be examined. This information can be provided by subject matter experts and non-experts. Such reviews are complementary as they provide different types of information. Experts can provide information about the accuracy of the content and possible problems the trainers might have using the system. Non-experts provide information about the presentation and style of the content and possible problems the *trainees* might have using the system. The types of questions that need to be answered by the expert and non-experts are shown in Table 6.4 (overleaf).

Table 6.4: Information required from expert and non-expert reviews (based on Tessmer, 1993)

	EXPERTS	NON-EXPERTS
Content Information	• Is the information complete? • Is the information accurate? • Is the information in a logical sequence?	• Is the information clear? • Is the information understandable? • Is the information easy or difficult to follow?
System Information	• Will trainers have problems using the system? • Will the system appeal to trainers? • Any problems with the visual display? • Any problems with the choice of input devices?	• Will trainees have any problems using the system? • Will the system appeal to trainees? • Any problems with the visual display? • Any problems with the choice of input devices?
Technical information	• Any problems with the visual quality? • Any problems with the audio quality? • Any problems with the tactile quality? • Any problems with the interactive quality?	• Any problems with the visual quality? • Any problems with the audio quality? • Any problems with the tactile quality? • Any problems with the interactive quality?

When choosing the appropriate experts and non-experts it is important to consider the type of outcome information that is required. Obviously there are many different types of experts (design, ergonomics, training etc.), and ideally it would be beneficial (but probably not practical) to use a variety of experts. However the most important thing for the success of a VET application is that it conveys the subject matter in an accurate and logical way, so it is important that at least one type of expert is a subject matter expert. As for the non-experts they have to represent as close as possible, the profile of the potential users (e.g. similar knowledge, skills, abilities, attitudes, personality etc.), so that their perception and views are similar to those of the potential users.

When considering appropriate review method for VET it may be more constructive to use face-to-face interviews. This is because a lot of the difficulties can be directly observed when the experts and non-experts are actually interacting with the application and these problems can be followed up with further questions. The kind of questions that need to be asked have already been suggested in Table 6.4 and the answers should be placed on some form of response sheet during the review.

As it is likely that only a small the number of people will be used for reviewing analysis can involve compiling the notes and eliminating comments that lead to impossible revisions (Tessmer, 1993). What is left can then be summarised into categories (e.g. the layout is too complicated, the text is unclear, etc.) and agreements and disagreements between the people in each group can be easily seen. The experts and non-experts can then be followed up if information is unclear (e.g. why is the layout too uncomplicated? why is the text unclear) and from these reviews a list can be compiled of all the changes that are suggested by the groups. The VE developer and the proposers then must decide from this list, what revisions should be made given the amount of time designated to the project.

6.6.1.4 Small group evaluations

Once the application is in its final stages of development small group evaluations can be used to provide some information on performance. In particular, as formative evaluation is still examining the design of the application, this evaluation considers how the design effects subsequent performance outcomes. In contrast to previous evaluation methods, the aim is to create a similar environment to the actual training environment. For example if the training is intended to be carried out from a home environment then the participants will be examined in a 'mock-up' of a home environment. Also unlike the previous methods, there is less direct interaction by the VE developer or the proposers with the participants as they carry out the training application. This is so that the VE developer and proposer can observe how the application might be used in the actual training situation.

The type of information required from the small group can be divided into four areas - effectiveness, efficiency, usability and appeal - as follows (Tessmer, 1993):
1) *Effectiveness* - The main reason for carrying out small group evaluations is to find out whether or not the application can fulfil the aims and objectives set out in the specification. This will involve determining this through some test or performance measure. For example if the VET application was to train a procedure, then if the participants are able to perform this procedure after training (assuming a pre-test indicated that they could not do this before), then the application may be considered effective. However if they fail then somewhere in the process there is a problem that has to be found.
2) *Efficiency* - Time is an important issue even for self-paced systems. Tessmer and Harris (1992) found that trainees will automatically set personal schedules and if the learning time exceeds this then trainees may abandon the instruction or pay less attention to it. Therefore it is important to assess if the trainees can master the subject in a reasonable time and also how time compares with the VET application and previous or other methods of training.
3) *Usability* - This is a very important issue when evaluating VET because of the problems that are already being identified with the current systems (see chapter two). A training application is a tool to enable learning of a subject and therefore should not require any specialised training before it is used. Therefore the application must be designed so that it can be easily implemented by the trainers and used by the trainees.

4) *Appeal* - As a new innovative technology VET already has some expected appeal, however given the current state of the technology, will the trainers and trainees still want to use it? and will they find it interesting?

6.6.1.5 Testing in the 'Field'

After revisions have been made from the information gathered during the small group evaluations the final formative evaluation should involve a 'field' test. This is to test the training application in a setting that is identical to the actual training environment to ensure that the method will transfer successfully from the laboratory to its intended purpose.

After formative evaluation which will have been carried out interactively throughout the development process, the application should be at its final version ready for implementation. This is when summative evaluation should be performed.

6.6.2 Summative Evaluation

Summative evaluation considers different levels of evaluation after the training has been conducted (see chapter five). The most widely known model is the 'four levels of evaluation' suggested by Kirkpatrick (1967; 1994) - Reactions level, Learning outcomes, Job behaviour and Results - these have already been discussed in chapter five and will be discussed with reference to VET as follows:

6.6.2.1 Reactions Level

The Reactions level is concerned with subjective feedback from the trainees. For the development of VET it is very important to consider what the trainees think about it. Generally research has shown that people have varying attitudes towards computers (Loyd and Gressard, 1995). It would be interesting to learn the attitudes towards VET as this is considered to be potentially more intuitive and engaging than other computer methods of training. Also the trainees can provide valuable feedback on the design features of the VR system and the VE in order to provide insight into how they interact with each other and the trainee to affect performance. There is currently no standard assessment tool for examining the features of VR and VEs therefore such a tool is suggested and used in the experimental part of this research (chapters seven and eight). A copy of this and its design can be found in Appendix II and is summarised as follows.

The assessment tool in its current form, consists of a number of checkpoints with rating scales. The tool is divided into two sections - assessment of the chosen VR system for the application and assessment of the actual design of the VET. Each section begins with an open question, "*What were your general impressions of the VR system/VE*?" This is to help the participant to think back on using the application. This question is followed by a list of categories of the attributes of the system and VE based on those discussed in chapter two.

The categories for assessing the VR system are shown in Table 6.5 and for assessing the VE these are shown in Table 6.6 (overleaf). Under each category is a number of specific statements relating to the information required from the assessment.

Table 6.5: Assessment of VR System Components

Categories	**Choices**	**Information required**
Type of VR System	• Desktop VR • Artificial Reality • Headset/immersive VR • Augmented Reality • Projected Reality • Other	- Appropriate choice of system - Relevant to application - Preferred type of system
Type of Visual Display	• Monitor • Headset • BOOM system • Glasses/Goggles • Large-Screen • CAVE • Other	- Choice of visual display - Size of display - Field-of-view - Clarity of display - Comfort - Preferred visual display

Type of Input Devices	• Touch-screen • Keyboard • Mouse/ • Spacemouse • Spaceball • Joystick • Wand • Glove • Movement platform • Other	- Choice of input device(s) - Ease of use - Ease of understanding - Use for movement around environment - Use for interaction with objects - Preferred input devices

Table 6.6: Assessment of design of VE

Categories	Information required
Layout of the screen	• position of different parts of the layout • colours used for the different parts • clarity of parts • size of parts • attractive/distracting
Visual appearance of VE	• realism • recognition • dimension • detail • colours
Use of sound	• amount • volume • clarity • realism • recognition • usefulness
Use of textual prompts	• understandable • useful/helpful • amount • clarity • size • colours
Use of auditory prompts	• understandable • useful/helpful • amount • clarity • volume
Use of pictorial prompts	• understandable • useful/helpful • amount • clarity

	• size • colours
Movement around the virtual environment	• amount of control • speed • accuracy • realism
Use of different viewpoints	• helpfulness/usefulness • understandable • amount
Interaction with objects	• selection • relevance • object behaviour • activation
Behaviour of objects	• realism • expectation • distraction
Sense of Presence	• involvement • feelings

The participant is asked to rate the statement on a seven-point scale from 'strongly agree' to 'strongly disagree'. At the end of each category the subjects rate it overall on a five-point scale from 'very satisfactory' to 'very unsatisfactory'.

6.6.2.2 Learning Outcomes

Evaluation of learning outcomes is especially important in VET as there is little evidence as yet to suggest that VET can enhance performance. The usual method of testing this is through using a control group that receives no training to measure whether, in the first instance, VET is an effective training method. Then to measure it against other methods of training in order to examine the potential cost/benefits of VET.

6.6.2.3 Job behaviour

The main aim of training is to affect a change in performance in the work environment, therefore the issue of transfer-of-training is very important. Again there are few VET evaluation studies in this area. Kozak et al (1993) and Kenyon and Afenya (1995) performed transfer-of-training designs using VEs for a basic pick-and-place task however revealed little information (see chapter four). Therefore this particular area also needs addressing as one of the potential applications of VET is for safety critical tasks where the transfer of training to the real world is essential..

6.6.2.4 Results

The results level of evaluation relates to the organisational needs. These needs are the driving force behind the technology. While there is on-going development in VET, generally the applications are in their early stages (and probably still involved in the formative evaluation level). Therefore, to date, this researcher has found no evidence of an organisation that has related VET with the fulfilment of an organisational training need.

6.6.2.5. Ultimate value

In another well-known model, Hamblin (1974), the final level of Kirkpatrick's model is divided into two - organisational outcomes and the ultimate value. The ultimate value is based on cost-effectiveness and cost benefits. Cost-effectiveness is the "*estimation of training costs and benefits in monetary terms*" and cost benefit is the "*examination of training in monetary terms compared to the benefits expressed in non-financial terms"* (Sanderson, 1995). There are different methods of calculating these however Dean and Whitlock (1983) suggest a list of the costs and benefits gained from using computer-based-training shown in Table 6.7.

Table 6.7 Costs and benefits from computer based training

(Dean and Whitlock, 1983)

Costs	***Benefits***
Start-up	**To students**
• people	• more effective use of time
• equipment	• available when needed
	• consistent presentation

On-going • course production including design authoring entry into computer correction review validation other media • course presentation running costs of equipment instructor time time of computer personnel • course updating	• reduced travel time and cost • social short-term requirements can be met • not away from place of work or home • practical training made more effective • may be available any time • home study may be possible **To training department** • reduced instructor time • easier student monitoring • accurate student monitoring • less classroom space needed • incentive to improve courses

These can be used to determine the cost benefits of VET and the cost-effectiveness by comparing the results with other methods of training. However as there are few 'working' VET applications and the VR technology is constantly changing, these analyses are difficult to perform as they involve much subjective judgement of *potential* costs and benefits. However a method for measuring non-financial costs and benefits is possibly through some form of assessment tool as suggested before (section 6.6.2.1)

6.7 Discussion

This chapter has considered the principles of the training development process and features which are particular to computers and training and applied them to the subject of VET. Many of the existing techniques employed in the development process are still relevant to VET development (e.g. training needs analysis, formative evaluation, summative evaluation, experimental design etc.) however it is possible to see that there are a number of areas which are specific to VEs and require further research. The first issue and an extremely important one is knowing when it is appropriate to consider VEs. This is the 'proposal' stage showed in Figure 6.2. There is limited research in this area and the lack of examples of ‘working’ applications provides little information for the potential VE user. In order to consider this a checklist of possible VE selection criteria was suggested. This is based on the training needs of industry for computer technology and informal interviews with many of the industrial case-studies described in chapter four. Secondly there are certain aspects detailed in the training specification (which is generated after training needs analysis) that require further development. For instance, the choice of the VR system can cause difficulties because there is a gap between the potential of the VR systems and the actual performance of VR systems. The choice of VR system affects the design of the VE, as the system will determine the type of VR software that can be used. This in turn is linked to the VE developer, that is, the choice of software will determine the skills required of the VE developer. Or depending on the VE developer’s experience the appropriate software will be purchased. The VE developer has a major influence at present over the design of the VE. There are no standards which exist that VE developers must adhere to, so the design of the VE is very much dependent on their knowledge, skills, abilities and attitudes. Also the translation process from specification to effective VET is still very much an uncertain area, especially as there is such little evidence of effective VET. Therefore in the first instance, there is a definite need to provide some evidence of the potential cost benefits of VET. Unfortunately the argument is circular – there is a need for evaluation before applications will be developed but there is a need for applications before they can be evaluated.

Given this situation the next chapter considers firstly the development of an application within the suggested framework and then explores three research questions:

(1) Is current desktop VR/VEs ready for application to training?

(2) Can VET effectively train basic psycho-motor skills?

(3) How does VET compare to other methods of training?

CHAPTER 7: Evaluation of Virtual Environment Training for a Basic Procedural Task - Experiment One

7.1 Introduction

This chapter explores virtual environment training (VET) for a basic procedural task. The previous chapters (one to six) have contributed to the research questions which were formulated at the end of chapter six and stated as follows:

(1) Is current desktop VR/VEs ready for application to training?

(2) Can VET effectively train basic psycho-motor skills?

(3) How does VET compare to other methods of training?

In this chapter a VET application is developed within the framework suggested in chapter six. In particular the evaluation stage will be closely examined using a controlled experiment as this is the area that has received little consideration in previous work regarding VET. Therefore the proposal stage of the application is outlined in section 7.2, followed by specification in section 7.3, building in section 7.4, evaluation in section 7.5 and ending with a discussion in section 7.6.

7.2 Proposal Stage

For reasons of availability of equipment and expertise, the tasks, which are part of computer maintenance, were chosen to be explored for this study. Through brainstorming sessions (see chapter five) with computer support experts a list of possible tasks was compiled. From task of replacing a network card in a computer was chosen as possibly benefiting from VE application.

7.2.1 Pilot Studies

Pilot studies involved informal surveys, interviews and direct observation in order to examine the task in more detail. Firstly, an informal survey of ten people who regularly use computers was conducted. Most of them did not know how to carry out this task, had never seen the inside of a computer and could not even guess how to do it. Three of this group were asked to carry out the task. They were given a computer, a network card and a screwdriver, but no instructions. They all had great difficulty and took about half an hour to complete this five minute task. Also, they were reluctant to touch anything inside the computer but also unknowingly mishandled a lot of the delicate parts. Therefore, to teach such a group how to replace a network card would require some form of training.

The usual methods of training this task are either reading from a manual or an expert demonstrating how to perform it. However each method has its advantages and disadvantages. Reading from a manual is certainly the cheapest form of training, as the information can be reproduced and given to as many people as required and the information is always consistent. However the trainee may have difficulty in interpreting some of the instructions because they are not familiar with the words or they are phrased in an obscure way. They may also have difficulty in visualising the task. Pictures are often used in manuals. However, these are two-dimensional and static and may not adequately convey the size and depth of the equipment, alternatively they may be too detailed making it confusing to see the salient points. Also they may not adequately show the actions required to make parts fit together easily or how much physical movement is required for each action. Manuals cannot provide the hands-on interaction which is necessary for trainees to overcome the reluctance of handling delicate equipment. They are a passive form of training and therefore, may not engage the trainees' interest. It may be also cumbersome finding answers (if they are there at all) in a manual. Also if the trainees are following instructions step-by-step, they do not have any feedback on whether they are doing it rightly or wrongly until maybe at the end when it hasn't worked. Then, to find out where any errors have been made they have to re-trace all their steps.

Another three people from the initial pilot survey were asked to carry out the task but this time they were given a manual to follow. They found many of the problems discussed above. In particular, it took time to interpret the meaning of the instructions especially the terms used and then translate them into the task they had to perform.

Many of these problems can be overcome by using an expert demonstrating the task. The trainee can actually see how the task is done and all the physical actions required. They can also ask questions if they do not understand and ask to see difficult parts of the process again. Then when the trainee practices with the real equipment, someone is there to provide instant feedback on whether they are doing it rightly or wrongly. However experts are expensive because not only are you paying for their time to train someone, you are also losing money by taking them away from their job. They are usually limited in supply, which means there are limitations on the number of trainees and training sessions you can have. Therefore using experts is not necessarily the most cost-effective method of training. Furthermore, being shown by an expert removes the self-pacing and personal exploration possible with other methods.

Another group of three people from the initial pilot survey were asked to carry out the task but first after watching an expert run through it. They all managed to successfully complete the task by asking a lot of questions while they were doing it. The expert also made sure that they were not mishandling any parts. However, they felt a little uncomfortable with being watched while they were doing the task and embarrassed by asking so many questions.

From these small studies, it would appear that a good method of training would allow the trainee to experience the cognitive and physical processes involved in the task in the real world. They should be allowed to continue at their own pace but be offered support and instant feedback. It had to be a method which was at least as good as the expert but was less expensive and capable of being reproduced and distributed to a large number of trainees. At this point, it was possible to see that VET had the potential for being valuable for training this task.

Considering it against the 'VE Selection Criteria' suggested in chapter six, the reasons why it was worth considering VET is shown in Table 7.1.

Table 7.1: Reasons for selecting VEs for the task

The Potential Trainees	• the trainees have different existing knowledge, skills & attitudes • the trainees already use computers as part of their job
The Task To Be Trained	• it is important for effective learning to experience the task in 3D through more then one sense (visual, aural, tactile) • it is important for effective learning to have some 'hands-on' interaction with the real equipment or process • it is important for effective learning to allow the trainee to practice a number of times on the real equipment or process • it is important for effective learning to allow the trainee to freely explore the real equipment or process • it is important for effective learning to allow the trainee to set their own pace • it is too expensive to use the real equipment or process for training of the task • the real equipment or process has limited accessibility or availability for training • monitoring how the trainee has performed the task is just as important as the end result • current methods of training are inadequate in some way • there are limited number of trainers of the task
The Working Environment	• there is limited access to the working environment for other reasons

Firstly, the people who would be recruited for the study would be those who already use computers as part of their jobs but were novices to the task. Therefore, having some experience with computers should mean that VET should not meet with too much resistance. They would also be selected from the

University environment and were likely to have different knowledge, skills, abilities and attitudes, so a method, which allowed them to continue at their own pace, would be potentially useful.

Secondly, the task is a basic procedural task, which requires the trainee to learn the procedure by rote. The most effective way of promoting procedural memory is through 'learning by doing' or practice of the equipment to be used (Wickens, 1987). It would also aid learning if the trainees could practice a number of times at their own pace. However it would be costly as some parts of the equipment may get damaged very easily and there are very few computers available to practice on. Also how the trainee performs the task is important, as parts of the task have to be handled with care. The current methods of training using manuals and experts have inadequacies as discussed before, and there is a limited number of computer support experts available to provide training.

Finally the computers are in continuous use in a busy working environment, therefore, it would be undesirable as well as unpopular, to allow trainees to practice in these environments. Therefore, VET has value to be explored for this task.

7.3 Specification of the VET application

Before the training specification for the VET application can be considered, the first step in the development process is to perform training needs analysis (see chapter five). From the results of these the VE specification can then be outlined.

7.3.1 Training needs analysis

As this task was developed for research there was obviously no need to perform an organisational needs analysis however a task analysis and person analysis were performed. The task was observed carried out by experts and non-experts and task and person analyses were used to identify the areas of difficulty and the potential skills required to perform the task.

- *Task analysis*

The task analysis (Stammers and Shepherd, 1994) produced is shown in Figure 7.1 (overleaf). The problems were as follows:

2.1 Locating the relevant network card. A computer can contain a number of different cards in a number of different places. The non-experts had a problem with recognising, which was the 'network' card to remove.

2.2.1 Locating the relevant fixing screw. The card is attached to a plate, which holds it in place in the computer. A 'fixing' screw secures this plate. However, around the location of the card there are a number of screws. The non-experts mistakenly unscrewed a number of these in an attempt to remove the card.

2.3 Taking out the old network card. There are many parts on the network card, which should not be handled, and some parts, which can easily break off. However, to take it out of its socket in the computer requires some force. In most cases the non-experts were initially careful with the card until eventually they realised that they would have to pull it out quite firmly. As they were unaware of the technique for doing this, they just grabbed the card and pulled, which unfortunately may have caused some damage to the brittle parts.

3.2 Placing new card in. A similar problem occurred as above when the non-experts were placing the new card in. They tended to try and force it while mishandling the parts on the card.

3.3.1 and 4.2.1 Relocating the fixing screws. Many people had a problem with finding the screws again once they had placed them down. Often they would drop them and they would be scattered around the floor.

4.1 Replacing the outer casing on computer. Finally, this seemed to cause a problem even to the experts. It appeared very difficult to get the holes for the screws to re-align. However the experts said that this was a common design problem in most computers, that is, the outer casing tends to warp after some use.

These appeared to be problems, which may be reduced by allowing the trainee to practice beforehand on a three-dimensional representation of the equipment like a VE. Therefore, VET would be used to address these problems.

- *Person analysis*

Considering the steps in the task outlined in Figure 7.1, the skills required by the trainee were basic recognition of parts, procedural knowledge and basic motor skills.

7.3.2 Specification for VE

A suggested format for a specification for the VE was outlined in chapter six and the sections, which are relevant to this study, are discussed as follows:

- *Description of the target training population*

The VET was to be aimed at non-experts. That is, people who did not already know how to perform this task. (A more detailed description of the participants in the study is given later in section 7.5.2.3)

- *Overall aim of project*

The overall aim of the project was to design, build and evaluate a VET application that effectively trained non-experts to perform the task of changing a network card in a computer.

- *Objectives of the VET application*

The objectives of the VET application (related to the research questions reported in the introduction to this chapter - section 7.1) were as follows:

1) To highlight features of the desktop VR/VE that may be beneficial to training applications;

2) To train non-experts with the VET application to effectively perform the task of changing a network card in a computer; and

3) To compare the value of VET over traditional methods of training.

- *Choice of VR system*

Choice of the VR system was based on the nature of the task. It was calculated by considering the cost of no training (i.e. damage to the equipment, approximately between £75 (new network card) and £2075 (cost of new computer and network card)) and the loss of revenue from not having the equipment active in use. Given the cost of VR systems (discussed in chapter two) it is possible to see that high-end VR cannot be justified (unless the task was absolutely vital in an organisation) therefore the choice clearly points to a low end option. The most appropriate system was desktop VR consisting of a Pentium 133 PC, a 14-inch monitor display, keyboard, mouse and spacemouse with Superscape VRT 4.00 software.

- *Choice of VE Developer*

The VE was built 'in-house' by VIRART because of their experience with developing desktop VR applications with Superscape software.

- *How the training is to be evaluated*

Given the objectives outlined above, in order to highlight features of VR and VEs, determine effectiveness of the application and compare it with another training method, it is necessary to measure the application against two other groups. One group would receive no training and the other would be trained with a more traditional method.

- *Timescale of the project*

Given the nature of the task, that is, it lasts about five minutes when fully competent and involves quite basic steps that can be easily remembered in a short time period, there would be little benefit in spending a large amount of time and money on development. So two months were allowed for designing, building and testing the VET before carrying out experimental trials.

- *The real environment*

As this application was developed for research the real environment was an experimental laboratory.

- *Content of the VE*

The specification must also include a description of at least the basic requirements of the VE (as discussed in chapter six).

Considering the analysis of the task illustrated in Figure 7.1, the description of preferred content is shown in Table 7.2.

Table 7.2: Basic Content of the Virtual Environment

Features	**Required in VE**
• Objects	- a table - a computer - a screwdriver - a network card
• Functionality	- table: no functionality - computer: parts to be dis-assembled according to user interactions - screwdriver: to be picked up and rotated - network card: to be removed from position and another to be picked up and placed in the correct position in the ‘appropriate’ ways
• User interactions	“pick up” screwdriver “unscrew” the appropriate screws “place down” screwdriver “lift off” the outer casing “place down” outer casing “unscrew” fixing screw on old network card “place down” fixing screws “take out” old network card “place down” old network card “pick up” new network card “place in” new network card “pick up” fixing screw “pick up” screwdriver “screw in” fixing screw on new network card “place down” screwdriver “pick up” outer casing

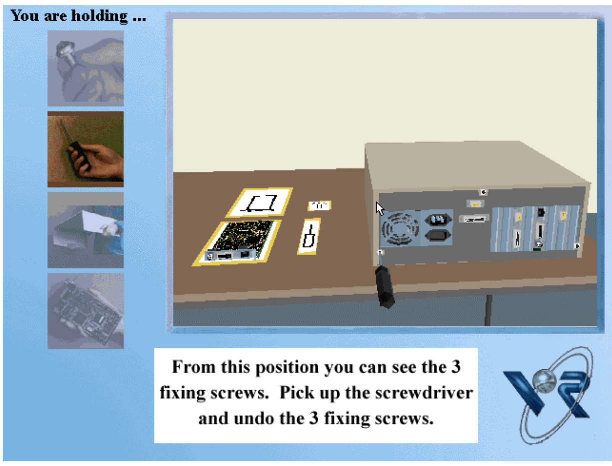
	"place on" outer casing "pick up" fixing screws "pick up" screwdriver "screw in" fixing screws
• Special requirements	- the trainee must be able to view the computer in as similar a way as possible to the real situation

7.4 Building the VET application

The building stage involved collaboration between the VE developer and this author (as the proposer of the application) through numerous review sessions. Issues included how to translate the specification detailed above in section 7.3, into an effective VET application given the requirements of the task and the constraints of the system (see also Eastgate et al, 1996). The application was referred to as the 'Netcard VET'.

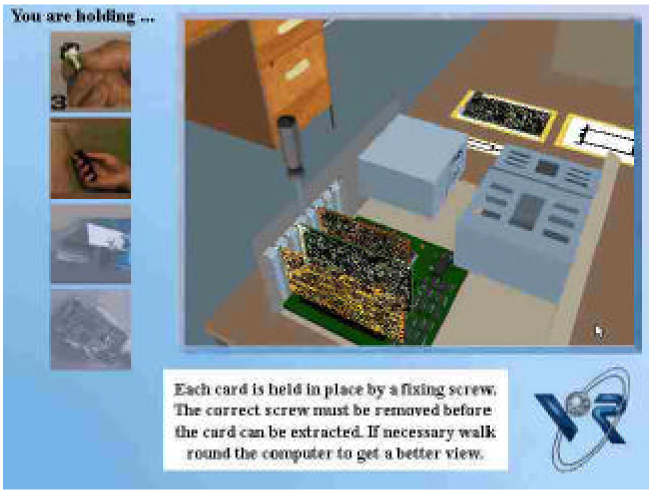
The first issue was how to present the relevant information in the VE to the user. This was tackled by laying out the screen in three parts as shown in Figure 7.2. The main part of the screen was a window into the VE. This was the only part of the screen the user interacted with. The other parts were to provide information. Below this window were textual prompts to provide the instructions as identified through the task analysis (see Figure 7.1). These automatically changed as the trainee completed the required step. If an error was made, a textual prompt would appear to inform the trainee.

Figure 7.2: Layout of the screen of the Netcard VET



A second issue was how to represent the user in the VE. Due to the constraints of the VR desktop system, the user had no embodiment representation (body or hand) in the VE. However, the task required objects to be "picked up", "placed down", "lifted off", "taken out" etc. While this was relatively straight forward - the metaphor of 'clicking' on the relevant object with the left button on the mouse replaced these actions - there was a problem with representing the psychological processes (psychological fidelity) correctly, as well as the physical processes (physical fidelity) (see chapter five Goldstein, 1993). For instance, in the real world, it is often necessary to have both hands free in order to perform an action like 'lifting off the outer casing of the computer' and it is obvious when your hands are not free because you feel the 'presence' of a screwdriver or screws without seeing them. However in the VE, where there are no virtual hands to be occupied or 'feel' anything, it becomes necessary to represent this in some obvious way. To not consider this problem would be training users incorrectly and may result in 'negative' transfer to the real world of the task, that is the training actually hinders performance of the real task (Goldstein, 1993). Therefore, 'picons' (picture icons) were used to provide feedback to the user on what they were currently holding in their non-visible 'virtual' hands. These were photographs of the various components of the task that the trainees were likely to have in their hands e.g. a screwdriver, screws, outer casing of the computer or the network card. In the case of the screws, as it is possible in the real world to have more than one screw in your hand, a counter was provided with the picon to inform the user of the number of screws they are 'virtually' holding. This is shown in Figure 7.3 (overleaf).

Figure 7.3: Use of Picons in Netcard VET



Another issue was how to allow the trainee to view the computer in a way which maps onto the real situation, given the input devices available to desktop VR. It was decided that the keyboard would be used to change viewpoints, and the spacemouse for movement and navigation. It was considered that two viewpoints - an 'average person standing' and 'average person bending' - would be adequate for completing the task as these were the two viewpoints generally used when completing the real task. The spacemouse was fixed to an average person's viewpoint as opposed to allowing complete freedom (that is, the ability to see the situation from any view e.g. bird's eye/worm's eye or ghost views). This is because of the usability problems experienced by initial users during the review sessions (discussed later). The spacemouse was found to be difficult to control and the many viewpoints tended to add further problems. It was very easy for the user to become 'lost' in the environment. In order to allow users to familiarise themselves with these functions, the task was modelled in a room. The first textual prompt then told the user to, “select the door to enter the room. Have a look around. Position yourself behind the computer” as shown in Figure 7.4.

Figure 7.4: Introductory Screen of Netcard VET

This was to allow the user time to get used to the system before reaching the task in order to relax them slowly into the technology.

7.5 Evaluation of the VET application

Different methods of evaluation (outlined in chapter six) were used to examine this application and to provide answers to the research questions at the beginning of this chapter. Firstly though, the application was examined through formative evaluation methods to ensure that the design and content of the Netcard VET were sufficient for the training. The second stage involved summative evaluation methods to examine the outcomes of the Netcard VET.

7.5.1 Formative Evaluation

Formative evaluation (Tessmer, 1993 - see chapter five) was used to examine in a systematic way the strengths and weaknesses of the design of the Netcard VET. The layers of formative evaluation used included self-evaluation, expert review and non-expert review and a small group or laboratory experiment as follows:

7.5.1.1 Self-evaluation

Between the author and the VE developer, the Netcard VET was examined using the following questions:

- Does the VE appear to fulfil the aims and objectives stated in the specification?
- Does it include all the 'content' listed in the specification?
- Are there any problems with the layout of the VE?
- Are there any problems with the overall appearance?
- Are there any problems with any of the prompts?
- Are there any problems with moving around the VE?
- Are there any problems with interacting with any of the objects?
- Do all the objects 'look' right?
- Do they move appropriately?
- Do they 'behave' appropriately?
- Are there any parts of the VE that need further discussion?
- Are there any technical problems with the system?
- Are there any problems using the system?

These resulted in a list of possible modifications which were prioritised according to the ones which were necessary and the ones which could be due to just personal preference. Modifications were made accordingly.

7.5.1.2 Expert and Non-expert review sessions

A number of review sessions were carried out with groups of experts and non-experts. The experts were computer support personnel who regularly perform the task of removing and installing network cards in computers. They were told that the application was to train non-experts to carry out the task and so they were to ensure that, firstly, they were able to carry out the task in the VE themselves and, secondly, whether the information provided was clear and accurate. This was done by allowing the expert to explore the VET application and then the author carried out an interview using the following questions (as suggested in chapter six):

- What are your initial impressions?
- Is the information complete/accurate/logical?
- Did you have any difficulties with using the system?
- Are there any problems with the visual display/input devices?
- What do you think of the quality of the display and interaction?
- What would you change about the application?

The experts were generally impressed by the application and felt that the information was sufficient, accurate and logical. They had little difficulty in using the system because they had experience with desktop VR and equally they had no problems with the visual display or input devices. The quality of the display and interaction were felt adequate for conveying the necessary information about the task and, overall, they felt that it was sufficient enough to be evaluated for effectiveness.

The non-experts were three people readily available to the author who had had no previous involvement in this study. They all had experience with computers, but not with VR systems, and said that they had never changed a network card in a computer before. They were told that the application was a self-pacing tool for training how to replace a network card and they were to explore the usability of the system and the presentation of the information. After explaining the different parts of the system, they were then left to examine the application after which they performed the task. This author directly observed the participants and then an interview was carried out using the same questions listed above.

General impressions were very positive and they all felt that the information was fairly clear, understandable and easy to follow (however there was a certain amount of 'participant bias' as they all eagerly agreed to pilot test the system). They found it easy to understand how to use the system and felt that the visual display was 'excellent'. However, they did say they initially found the spacemouse input device difficult to use. As mentioned above, the participants were given complete freedom over which angles they wanted to see the equipment in the VE. The spacemouse offers this capability by translating pressure, applied by the participant's hand on its surface, into movement on the vertical and horizontal planes at the same time. This allows the participant to see the situation from great heights (bird's eye view), lows (worm's eye view) or even impossible views (ghost view). However, the ability to understand this 'new' form of interaction with a computer requires some time and practice. The participants had difficulty in controlling these viewpoints and tended to either 'spin' out of control or get themselves into positions where they were lost in the VE and could not get back to the task. From discussions with these participants, it was decided that to reduce the problem the spacemouse would be fixed to the view of an 'average' person and in order to change views quickly and easily the keyboard would be used. As stated before, two viewpoints, as agreed by the experts and non-experts, were felt adequate for completing the task - an average person standing and an average person bending. These modifications appeared to provide the participants better control of the situation. Also, further suggestions made by the non-experts led to a few modifications made on the phrasing of the textual prompts. The revised versions of the application were continually reviewed by the same group of people, until everyone felt that it was sufficient enough to be evaluated for effectiveness and transfer.

7.5.1.3 Small group or laboratory experiment

This will be discussed in more detail next. During a laboratory experiment, direct observation of the usability of the system was used to note any difficulties, further supported by an assessment tool examining the features of VR and VE (see Appendix II). The results are discussed later in section 7.5.2.7.

7.5.2 Summative Evaluation

Summative evaluation considers the outcomes of the training. As this application was developed for research, the only levels of the four levels of evaluating outcomes (Kirkpatrick, 1994 - see chapter five) that could be examined were level

1: Reactions and level 2: Learning outcomes; as the other two levels specifically relate to the 'on-the-job' situation and organisational strategy. However, an attempt at cost-benefit analysis and cost-effectiveness are also provided to give an indication of the potential financial value of the technology. To gather information for these evaluations an experiment was conducted.

7.5.2.1 Research Objectives

The objectives of the VET application (stated in the specification outlined above 7.3.2) were as follows:

1) To highlight features of the desktop VR/VE that may be beneficial to training applications;

2) To train non-experts with the VET application to effectively perform the task of changing a network card in a computer; and

3) To compare the value of VET over traditional methods of training.

In order to fulfil the first two objectives the application would need to show some transfer-of-training to the real task and the last objective would need a comparison of the performance and opinions of trainees of a VET and another method to be made. To do this, the application was measured against participants that received no training to show the benefits of VET and also measured against participants trained by an alternative method, in order to prove the value of using VET. As currently this task is usually trained by using some form of demonstration (i.e. video or expert) the alternative method chosen to compare with VET, was a training video.

In order to measure effectiveness and transfer, performance data was collected. This was the performance of the participants after training, on completing the task in the real situation in terms of time taken and errors made. Also as relatively little is known about the effects on the participant of a new technology like VET, it is also important to gain attitude and opinion data from the trainees through questionnaires. The experimental hypotheses are therefore as follows.

7.5.2.2 Experimental Hypotheses

H_1: Participants trained by the VET will perform the task faster than participants which have received no training.

H_2: Participants trained by the VET will perform the task faster than participants who have been trained by the video.

H_3: Participants trained by the VET will make less errors than participants who have received no training.

H_4: Participants trained by the VET will make less errors than participants who have been trained by the video.

H_5: There will be a difference in the opinions of the participants effected by the training method experienced

7.5.2.3 Participants

Thirty people participated in the experiment with an equal divide of male and female participants. Ages ranged from 20 - 40 years old, with the majority (20/30) being between 20-25 years old. They were either self-selected as respondents to posters placed around the University or recruited by direct contact (i.e. engaged while in the coffee bar). They were a combination of postgraduate students (15/30), researchers (10/30) and University staff (5/30). They were required for one session lasting about 40 - 50 minutes in total, and were paid £4 for their participation. People were selected if they had never changed a network card in a computer before but had experience with computers.

7.5.2.4 Independent and Dependent Variables

The independent variable was 'training' and this was measured at three levels: (i) training with the VET application; (ii) training with a video; and (iii) a control group which received no training. The dependent variables were time, errors and opinions.

7.5.2.5 Measures

- Time was measured by a digital stop watch.

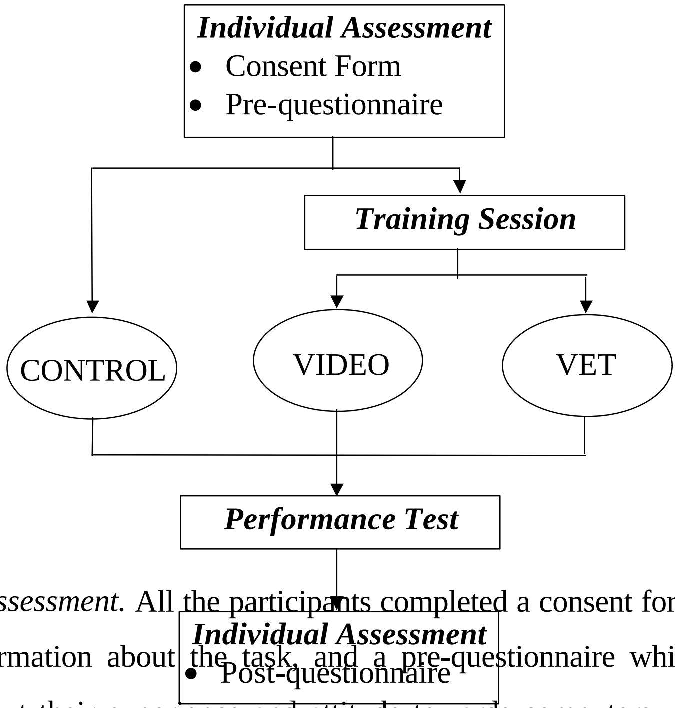
- Errors in performance of the task in the real world was measured through direct observation and video analysis. An error was considered to be any deviation in the steps of the task outlined by the task analysis (see Figure 7.1).
- Opinions were measured through questionnaires consisting of open questions, multiple choice and rating scales at the beginning and at the end of the experiment.

Further information about the design of these measures and copies of each can be found in Appendix (III).

7.5.2.6 Design and Procedure of experiment

A between-subjects design was applied as through pilot studies it was shown that the task could only be trained once. The experiment had three stages of data collection - individual assessment, training and performance test. These are illustrated in Figure 7.5 (overleaf).

Figure 7.5: Procedure of experiment one



(i) Individual Assessment. All the participants completed a consent form providing them with information about the task, and a pre-questionnaire which collected information about their experience and attitude towards computers. This author reiterated what was said in the consent form, i.e. they were told that they would be trained to replace a network card in a computer. Then after the training and the performance test, a post-questionnaire was administered and if they had been trained by the VET application, a further questionnaire on the features of the VET application was completed. Copies of all of these questionnaires can be found in Appendix (III).

(ii) Training. When the participants had completed the first set of questionnaires the training stage was carried out immediately after. The participants were assigned randomly to the control group, Video group or VET group but an equal number of male and female participants were maintained in each group. The control group received no training but instead proceeded straight to the performance test

The Video group were trained using a video of an expert demonstrating the task. The video allowed the demonstration to be consistent to all the participants in that group. They were allowed to sit down where they were the most comfortable to view the television screen which was a standard 21-inch colour screen. The video lasted approximately five minutes and the participants viewed it through once. No participant wanted to see it again.

The VET group were trained using the VET application. Firstly the components of the system were explained including instructions on how and when to use the input devices. Then the participants were given control and told to follow the prompts on the screen. From this point the participants were allowed to proceed at their own pace and to ask questions if required.

VIRART produced both training methods - video and VET - therefore exactly the same instructions were provided to both groups. These instructions were in the form of discrete steps outlined by the task analysis (see Figure 7.1). The training video consisted of a ‘voice over’ slowly explaining each step of the task as an expert performed it. The steps were set at a steady pace with particular care focussed on areas which were highlighted by the task analysis as providing potential difficulty. In contrast, the participants of the VET application set their own pace. The Netcard VET provided the same instructions through a series of textual prompts which were automatically delivered only when the participant completed each one correctly. If the action was incorrect, e.g. the wrong fixing screw was chosen, the wrong card was chosen etc., then a textual prompt would appear informing the participant that they had made an error. In this way the participant was also aware of some of the potential mistakes that they could make.

Information was gathered about each training method through open questions and rating scales in the pre- and post-questionnaires, direct observation and video recording through the process.

(iii) Performance Test. Immediately after training, all the participants carried out the same performance test. The participants were led into the experiment room and were asked to stand behind the experimental table. The table was set out with five boxes marked out in black tape on its surface. Each box was labelled with what it contained or should contain and the relevant piece of equipment was placed in the boxes at the start of each experiment. The positions of the equipment were confined in this manner so that the task could be set up exactly the same for every participant and the task could be performed in the same way. Also pilot runs had showed that without controlling where the participants placed the screwdriver and screws, some had the tendency to lose them which affected performance time. Instructions were also provided on the table in case the participants needed them.

The participants were told that they were to complete the task of changing a network card using the equipment on the table, as quickly and as accurately as possible, while being timed and videoed. Also any errors they made would be pointed out as soon as they moved on to the next step and they would have to return to correct it. (This was so that errors would incur time penalties and the classical SATO - Speed/Accuracy Trade-Off problem discussed by Drury (1995) - would not occur.) The task was completed when everything had been returned to their relevant boxes. The participants only performed the task once as it was found through pilot tests that after the first time there were no difficulties in performance the second time.

7.5.2.7 Results

The results will be discussed using the three stages of data collection outlined in the procedure. The groups will be referred to by their method of training i.e. control, Video and VET - throughout this section. A more detailed breakdown of the results can be found in Appendix (III) but is briefly outlined as follows:

(i) Individual Assessment. A summary of the subject profile is shown in Table 7.3 (overleaf). As Table 7.3 shows, the majority of the participants were between 20 -

25 years old (20/30) with an equal divide of male and female in each group. They had a lot of computing experience with just over a half (16/30) of the participants using computers everyday (except more of those in the video group used computers at least once a week (6/10)).

For the whole group alot of this experience was in word processing (23/30) with some experience of games (17/30) and spreadsheets (16/30) and no experience of programming (17/30) and maintenance and repair (26/30).

Table 7.3: Profile of subjects per group in experiment one

	CONTROL	**VIDEO**	**VET**	**TOTAL**
Age	20 - 25 (8)	20 - 25 (7)	20 - 25 (5)	20 - 25 (20/30)
Gender	male (5) female (5)	male (5) female (5)	male (5) female (5)	male (15/30) female (15/30)
Use of computers	everyday	once a week	everyday	everyday (16/30)
• Games	some	some	some	some (17/30)
• Word processing	alot	alot	alot	alot (23/30)
• Spreadsheets	some	some	some	some (16/30)
• Design	none	none	some	none (16/30)
• Databases	alot	some	some	some (19/30)
• Internet	alot	some	some	alot (13/30)
• Programming	none	none	none	none (17/30)
• Maintenance & Repair	none	none	none	none (26/30)
Virtual Reality	yes (10/10)	yes (10/10)	yes (10/10)	yes (30/30)
• newspapers/magazine	9	7	6	22/30
• TV, film, video	8	9	10	27/30
• observed use	4	4	4	12/30
• experienced use	6	4	3	13/30
• worn headset	6	4	5	15/30
• played game	4	1	2	7/30

• other application	4	3	3	10/30
Computer Attitude Scale	High (132.2)	High (131.5)	High (132.2)	High 132/140
Seen inside computer Seen a network card Experience of task	yes (5) yes (2) none	yes (4) yes (2) none	yes (5) yes (2) none	yes 14/30 yes 6/30 none 0/30

However the control group appeared to have ‘a lot’ more experience of databases (5/10) and the internet (6/10) than the other two groups and the VET group appeared to have ‘some’ experience (4/10) of design compared to the other two groups. As for VR, all the participants had heard of the term, the majority of the participants had seen examples on television, video or film (27/30) and had read about it in newspapers and magazines (22/30). Of the other types of experience the control group appeared to have had more experience of using VR systems (6/10), wearing a headset (6/10), playing games (4/10) and other applications (4/10) then the other groups. Given the type of experience, most of the participants had a good understanding of VR with the words ‘real life or real world’(21/30), ‘computer-generated’ (18/30), ‘interactive’ (6/30) and ‘simulations’ (6/30) being the most frequent words used in their definitions as revealed through content analysis (Bainbridge and Sanderson, 1995 - see Appendix III).

Applying the Computer Attitude Scale (Loyd and Gressard, 1995) all the participants had a high positive attitude towards computers with few differences between the groups (Control = 132.2/140; Video = 131.5/140; VET = 132.2/140). Further subscores of anxiety, confidence, liking and usefulness, equally resulted in few differences.

Finally with regards to the task, none of the participants had had any experience of this particular task beforehand; just about a half in each group (14/30) had seen the inside of a computer and a fifth (6/30) knew what a network card looked like.

Generally, the three groups were fairly equally matched in terms of computing experience and attitude and all the participants required some form of instruction to complete the task.

(ii)Training. Before training the participants were required to rate a list of words from strongly agree to strongly disagree in order to provide some base line of their expectations. This list consisted of 30 words - positive, negative and descriptive - associated with training e.g. effective, stressful, self-pacing. (Further information can be found in Appendix III). This same list was administered after training to indicate any differences from their expectations and to highlight any differences in opinions between the training methods. Likert scales were used in order to obtain parametric data and *t*-tests were applied to each word from the video and VET group responses. While before training there were no significant differences in the opinions between and within these two groups, after training some significance was found. These are shown in Table 7.4:

Table 7.4: Significant differences in opinions between the groups

Factors	Significance
interactive	t = 5.072; df = 18; p<0.05; one-tailed
self-pacing	t = 3.773; df = 18; p<0.05; one-tailed
interesting	t = 3.893; df = 18; p<0.05; one-tailed
enjoyable	t = 2.765; df = 18; p<0.05; one-tailed
incomprehensible	t = 3.151; df = 18; p<0.05; one-tailed
too hard	t = 2.496; df = 18; p<0.05; one-tailed
boring	t = -3.431; df = 18; p<0.05; one-tailed
uninteresting	t = -2.678; df = 18; p<0.05; one-tailed

From these results it is possible to see that the VET group found the VET more ‘interactive’, ‘self-pacing’, ‘interesting’ and ‘enjoyable’, but also more ‘incomprehensible’ and ‘too hard’ compared to the participants of the video group.

The video group however found the video more ‘boring ‘ and ‘uninteresting’ compared to the VET group. These opinions can be confirmed through observation of the behaviour of the participants during training. Observation analysis included noting the number and type of questions and comments made in terms of the system, the VE, the instruction or any other features of the training and identifying any difficulties in specific steps of the instruction (copies of the observation forms used for analysis during training can be found in Appendix III.)

The Video group were very much passive viewers and the total training time was consistently about six minutes. They generally remained still while watching the video and appeared to understand what was required of them as no questions were asked before, during or after the training, although the opportunity was given. In contrast, participants in the VET group took different amounts of time to complete the training, between 6.03 and 15.33 minutes (mean = 10.53 minutes, SD = 2.63).

Also they asked a number of questions, mostly about how to use the spacemouse. The majority of the participants had never used a spacemouse before and two participants were quite reluctant to take control of the device themselves. The majority (9/10) had problems initially controlling the device and questions like ‘where am I now?’ were frequently asked when they lost control and ended up in a difficult situation. Four participants made the comment that it required a lot of concentration (which could have had an affect on their ability to remember the steps of the task). Also switching from one input device to another (the spacemouse to mouse to keyboard) was commented on as distracting. Two participants frequently asked questions about the viewpoints. A ‘bending’ viewpoint was given so that that participants could get a closer look at the equipment. However to carry out some of the task it was better to be in the ‘standing’ viewpoint and in most cases the experimenter had to advise the participants when they could not understand how to do something. (In the real world of course, it would be obvious that you would have to bend or stand upright without requiring much thought but this was not instinctive in the VET). On the whole though, all the participants were able to follow the instructions

These observations were further supported by the qualitative data from the questionnaires. The general impressions of the video group was that the video was *"... clear, uncomplicated, unfussy. Simply told you what needed to be done and how to do it*", "*Straightforward - easy to follow*". However they also thought that it was "*... a little bit too simplistic*", "*... quite slow*". The majority of the video group (7/10) felt that they had benefited from the training and reasons given were mostly because they were then able to perform the task. Of the remaining three responses, two of the participants said that they had already carried out similar jobs and the other participant said that they would have preferred to have seen the real equipment. Suggested improvements to the application were mainly better pictures of the equipment.

The comments made by the VET group were generally positive - "*it was nice to know what you were getting yourself into*", "*straightforward and interesting*", "*concise and left me in no doubt as to what to do and what to recognize in the real hardware*". However one participant felt that the task *"... could have been done equally well with written instructions*", and another said it was "*... good but would have been better if didn't have to switch between views i.e. F1,F2 ...*". The majority of the VET group (9/10) said that they had benefited mainly because it showed them exactly what the task would *look* like.

Some of the comments are as follows: *"could run through the task, didn't matter if you made mistake because you weren't thinking about breaking anything"; "showed me what to expect and allowed me to identify all the elements/equipment I was going to use"; "when confronted with the real hardware I already knew what to look for";* and *"changing a card is something I have not done before and I thought it was complicated but I now know differently"*.

As for whether all the necessary information was provided, in the main (9/10) participants felt it was. In particular a participant said “*additional information boxes about feelings I would experience e.g. with initial force, remove card - meant when I did task I wasn't afraid to give the card a 'tug'* “. A participant that did not feel all the information was there said that you “*didn't have to think about which way the card went in or out. As soon as you picked it up using the mouse it was inserted correctly*”. Suggested improvements were to show that the card had to be orientated.

Evaluation of specific features of the VET application was carried out using the VR assessment tool discussed (see Appendix II) consisting of rating scales. The results have been summarised and combined to produce Tables 7.5 and 7.6 and a detailed breakdown can be found in Appendix III.

As shown in Table 7.5, overall the participants found the main features of the VR system ‘satisfactory’. However some would have liked a headset (4/10) (although only two participants felt that it would have been beneficial) and preferred a joystick (4/10) or an on-screen device (5/10) for movement.

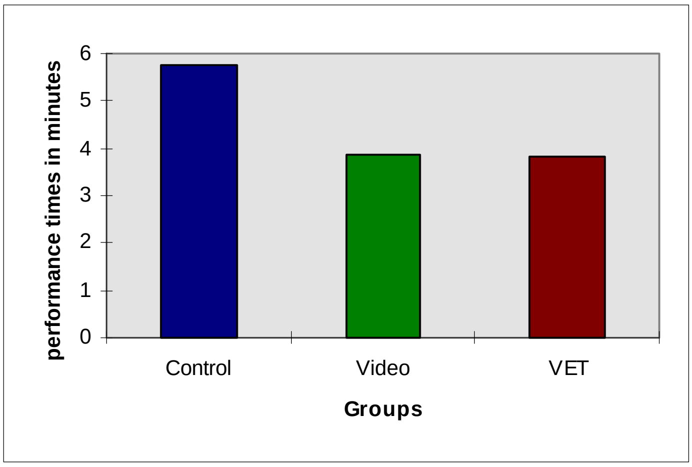
These results were supported by the qualitative information gathered. Comments on impressions of the system were generally positive with the words ‘good’ and ‘enjoyable’ being used frequently. However some participants made comments about the difficulty they had with the spacemouse which are summarised by the following participants comment: *“Although I understood how the input device moved me round the environment, it was very sensitive and so difficult to be accurate. It was frustrating to use as I kept going too far. However I do think with even a short time (10-15 min.) it would have become a lot easier to use...”*

As for the design of the VE, as Table 7.6 (overleaf) shows the majority of the participants were satisfied with the main features of the VE. The only areas that some participants felt were not ‘satisfactory’ were in the ‘use of sound’ where over half of the participants (6/10) requested more sound in the VE. Regarding the ‘pictorial prompts’, about a fifth (2/10) felt that they were not useful or helpful. As for movement around the VE, a fifth (2/10) felt that the movement was not right (too fast), a quarter (2.5/10) did not find it easy to move around and about a third, did not feel in control (3/10) or able to achieve the right viewing angle (3.5/10).

Considering the comments about the input devices made above, these feelings can be attributed to the spacemouse. Almost a third of the participants (3/10) felt that the behaviour of the objects were distracting and finally only one participant had a low rating of presence when applying the Witmer and Singer (1994) presence questionnaire. This participant also made the following comment when asked to give their general impressions of the VE, "*I don't like it that much - I would have preferred another type of training method*". Generally though the majority of the participants were satisfied with the system and the VE.

(iii) Performance Test. All the participants were able to complete the task. The times taken by the three groups are shown in Figure 7.6, as follows:

Figure 7.6: Performance Times of each group



The means and standard deviations of the groups were as follows:

- Control: mean = 5.75 mins. (345s; SD = 168.61)
- Video: mean = 3.87 mins. (232s; SD = 66.05)
- VET: mean = 3.84 mins. (230s; SD = 107.96)

Examination of the data shows large differences in variance between the three groups (F_{max} = 5.69; $F_{crit}(3;9;p<0.05)=3.86$; $F_{max} > F_{crit}$), moderately skewed in a positive direction ($z > 1.96$), so they do not meet the assumptions required for an analysis of variance (ANOVA). A transformation of the data was carried out, which corrected the variability and allowed for a one-way ANOVA to be performed. This revealed no significant differences between the performance times of the three groups (F =3.111; df = 2,27; p>0.05).

However an application of the transfer of training equation (Patrick, 1992) as below:

$$\% \text{ savings} = \frac{B_2 - B_1}{B_2} \times 100$$

where B_1 = training group;
B_2 = control group

shows that the video group made savings of 34.3% on time and the VET group made savings of 35.3% on time, showing that the training methods had some value over the control group.

Even though the training times appeared to imply that few errors were being made, observation analysis and the comments made about the task in the questionnaires revealed that there were aspects of the task that the groups found difficult. One of the main differences observed between the participants regardless of group was the use of the screwdriver to undo screws. Some participants appeared to have real difficulties in controlling the screwdriver which resulted in dropping screws, regardless of which group the participant belonged to. This obviously had implications for the time taken to perform the task. Also every participant had difficulty in getting the outer casing back on the computer. This however according to the computer support experts is a common problem with most computers, their outer casings tend to warp out of shape. The problem unfortunately cannot be trained as every computer requires its own individual technique.

Instructions were on the table in case they were needed. Obviously all the participants in the control group used them as they were unaware of how to perform the task. Some of the control were quite nervous as it was difficult for them to know if they were interpreting the instructions correctly. This was highlighted by their responses to the question of their initial impressions of the task, as follows: *"I found it quite daunting - I am just not at all practical and can barely hammer in a nail"; "More tricky and fiddly than I expected"; and "Not as straight forward as expected, primarily because I didn't realise there were three cards and although it was easy to identify the correct one I was worried in case I was wrong."* In contrast the video and VET groups used words like "*straightforward..., simple... and easy*" to describe the task.

Some of the control asked for confirmation on parts of the task and these questions were only answered on a 'yes/no' basis so that they had to make choices which could be noted down. There were mainly three questions asked: (i) Is this the right card? (6/10); (ii) Have I taken the right fixing screw out? (4/10); and (iii) Can I take the card out now? (8/10). Generally the participants who were trained began the experiment immediately and did not ask any questions and tended to perform the steps without any further reference to the instructions.

In terms of particular areas of difficulty with the task, the control group, as expected, commented on problems that were identified by the task analysis: "*recognising the relevant card, locating the relevant fixing screw, taking the old card out, placing the new one in and placing the outer casing back on*". The control group also had problems with relating to the instructions provided. However, these were written exactly as expressed in the video and the VET, implying that the use of just written instructions was not sufficient and that reinforcing them by some visual means appeared to be better. The video group (except for one participant) also encountered similar problems, implying the video was not successful in training for these problem areas. Of the VET group, six participants said they encountered no difficulties and the remaining four said they had problems with replacing the fixing screws, the outer casing and removing the card. Therefore, although there was very little difference in mean performance times of the video and VET groups, they did experience difficulties and the VET group reported less difficulties. This suggests that the time measures may not have been sensitive enough to highlight the errors that were being made.

7.6 Discussion

This chapter has described the development process of a VET application through the stages of designing, building and evaluation (as suggested in chapter six). In particular the objectives of the VET (linked to the research questions) were as follows:

1) To highlight features of the desktop VR/VE that may be beneficial to training applications;

2) To train non-experts with the VET application to effectively perform the task of changing a network card in a computer; and

3) To compare the value of VET over traditional methods of training.

Before these are discussed, firstly the results will be considered under the evaluation levels of 'reactions', 'learning outcomes' and 'cost- benefits/cost-effectiveness' analysis as follows:

Level 1: Reactions Level

Generally the majority of both groups felt that they benefited from the training and all the necessary information was provided. A comparison of the results of the two groups revealed a difference in their opinions about the method they were trained by. The VET group naturally found their method more 'interactive' and 'self-pacing' than the video group, however they also found the VET more 'enjoyable', 'interesting' but 'incomprehensible' and 'too hard' compared to the video group. However, this is not to say that the VET was incomprehensible or too hard by direct comparison, as the groups had different subjects. If the comments by the video group are considered, they show that the subjects found the training "*clear, uncomplicated, unfussy..., a little bit too simplistic..., quite slow*". In comparison the VET was less comprehensible and harder. They also felt that the video was more 'uninteresting' and 'boring' than the VET.

In the main the VET group were very positive about the VET but observation results and information gathered through the questionnaires confirmed usability difficulties with the spacemouse by some participants. Generally they found the spacemouse difficult to control and their concentration during the training was mainly on using the system rather than learning the task. The spacemouse is a standard input device for desktop VR, therefore further research is required in establishing the amount of time required for pre-training of input devices before training of the task begins. The majority of the participants liked the features of the VET except for one participant who just did not enjoy it and would have preferred another method of training.

Therefore the reactions of the participants were different but as the responses were not a direct comparison by the participants themselves it was not possible to determine which method was preferable, therefore it would be of interest to allow a group of participants to experience both methods so that they can directly compare the similarities and differences.

Level 2: Learning Outcomes

The performance times of the video and VET groups did not show a significant difference from the control group but did show some time savings (34.3% and 35.3% respectively). Also the trained groups did not need to ask any questions during the training and were generally more confident about doing the task then the control group. The lack of evidence of significance in the results can be attributed to the nature of the task and the measures used. The task was basic but not immediately obvious. However as the control group were able to complete it with just instructions and with little error, VET was unable to show much more learning benefits over this. However the control group did express some difficulties in the task but these did not provide long enough time penalties to be picked up by the time measures. Overall the difficulties encountered by all the groups were generally to do with motor skill ability which was difficult to convey with the desktop VR system used. This was further confounded by the difficulties the participants encountered with the spacemouse input device. Therefore in the case of psycho-motor skills where there is more emphasis on 'motor' skills rather than 'psychological' skills, VR systems are fairly limited. Thus reaffirming the findings of Kozak et al (1993) and Kenyon and Afenya, (1995) that VEs may be of more use to tasks which involve more psychological processes than motor processes.

Level 3: Cost-benefits and cost effectiveness

The cost to an organisation of developing these training methods are different in terms of time and money and potential value. A summary of the costs is shown in Table 7.7 (overleaf).

Based on quotes by a number of video production companies contacted locally, assuming the organisation can provide a detailed 'storyboard' and an expert to perform the task, the approximate costs per day are as follows: shooting of the video including two people - £550; travel expenses - £250 a day or £0.35p per mile; 30 minute tape - £15.

Table 7.7. Cost benefits and cost effectiveness

Cost/benefits	**Professional Training video**	**VET application**
Development		
time money	~ 2 days ~ £2,000 +	~ 2 months £14,000 - £32,000
Running costs		
equipment personnel	£400 – television/video	£1500 - Pentium PC £100 - software Computer support

Benefits		
	• Available when needed • Consistent presentation • Reduced instructor time • Familiar format	• Available when needed • Variety of presentation styles • Reduced instructor time • Novel approach • Self-pacing • Interactive method

Therefore, for a 5-10 minute video, it would take on average two days work from shooting to final copy at a development cost of approximately £2000 upwards. In comparison, the development time and money of a VET varies considerably based on the VE developer. Again assuming the organisation can provide a detailed storyboard and an expert to perform the task, an experienced VE developer would need about two months in order to specify, build and test the VET before implementation. Rates for such work are generally negotiable, however academic research groups such as VIRART begin at £350 a day while commercial developers such as Superscape begin at £800 a day. Therefore, the development cost of this VET could be anywhere between £14,000 and £32,000.

As for running costs, the price of a 21” television with Nicam digital stereo and a video machine to run a quality training video begins at £400. Also as no specialist knowledge is required to play a video, running costs do not include a technician. The VET requires a computer and currently the minimum specification of a new computer is a Pentium PC at approximately £1500 and software to visualise the VE at approximately £100. Further to this cost, some computer support is required for setting-up the equipment and ensuring that there are no problems when it is running.

In terms of benefits, both training methods are available when required however a video has consistent presentation so that every trainee receives the same quality of training, the use of a video minimises the time an instructor is required and trainees are used to the video format and so can concentrate on what is being taught. Alternatively, the VET also reduces the time an instructor is required as it is a self-pacing medium i.e. it allows the trainee to ‘self-teach’ at an appropriate pace. Finally, it also uses a variety of presentation styles which allows the trainee to be interactive and can sustain the trainee’s interest in the training (but possibly distract from the subject matter). Quantifying these benefits is difficult and largely subjective. It depends on the most appropriate way of learning the subject and the importance of the training.

Therefore returning to the objectives of this VET.

(1) Features of the desktop VR/VE that may be beneficial to training applications

The VET application was able to provide experience through a fairly accurate visual representation of the task. However it was unable to provide experience of a physical representation. However the nature of the task was such that the control group were able to complete it using a list of instructions and therefore further benefits of the VET application over no training were difficult to show.

2) Effective training of non-experts to perform the task of changing a network card in a computer

The VET application was successful in training the non-experts as they were able to complete the task with no errors and without the use of further instructions with the minimal amount of physical or mental strain

3) Comparison of the value of VET over a traditional method of training

The value of the VET over the video appeared to be greater in non-financial terms but unable to compete in financial terms. The VET appeared to create a more positive impression than the video and thus in motivation of the trainees the VET appears superior. However the development time and the training time exceed those required for a Professional video and so in terms of cost-effectiveness VET has still some way to go.

The results of this first experiment provided some insight into the development and evaluation process of VET for a basic procedural task but in order to address some of the issues highlighted further, a second experiment was performed. This is outlined in the next chapter - chapter eight - and the task chosen again requires psycho-motor skills but emphasises the ‘psychological’ skills more so than motor skills. Also the participants were required to experience both methods of training so that they could directly compare the strengths and weaknesses of each.

CHAPTER 8: Evaluation of VET for assembly procedural training: Second Experiment

8.1 Introduction

This chapter outlines a second experiment which examines VET. The results of the first experiment (see chapter seven) highlighted some important issues related to the use of VET; however the nature of the task trained probably obscured significant differences in training effectiveness. Therefore a second study examined a different task requiring psycho-motor skills, but with more emphasis on psychological skills and allowing participants to experience different types of training for a more direct comparison to be made. Again the VET application was developed within the framework suggested in chapter six, so the stages of proposal, specification, building and in particular structured evaluation will be discussed in sections 8.2, 8.3. 8.4 and 8.5 respectively. The chapter will end with a discussion of the main points.

8.2 Proposal Stage

As already discussed at the end of chapter seven, the task of changing a network card had a number of shortcomings for exploring VET applications. Mainly the task was too quick and simple and therefore could only be trained once and the errors which could be made were minimal and not well reflected in the time measures. Differences in times for subsequent task performance were partly caused by participants' varying abilities in using the screwdriver, which is something which is difficult to train with current VR systems not using force/tactile feedback. Also, one of the most difficult parts of the task was replacing the outer casing on the computer, which was a design fault and cannot be trained as every outer-casing is different on every computer.

Therefore a new task was considered which had the advantages of the first task and addressed its disadvantages. This should be a task requiring the user to follow basic procedures which could be taught in a short time, but would prove difficult, if not impossible, to carry out without training. A number of options were considered. The decision was between a task which already existed or designing a generic task which required basic procedural training. Again the advantages and disadvantages of both were examined but it was decided to minimize any task-specific components and design a generic application, as it would offer greater control over the variables in the experiment.

In order to eliminate the use of tools (actual and virtual), as varying manipulative abilities can have misleading time implications, the generic application used 'Lego' building blocks. This is because many people could be sampled who already have skills using 'Lego' and for those who have never used it before it is not difficult to fit the pieces together.

A number of different Lego models were examined to find out if they were obvious to build without instructions and how long they would eventually take to build. The model chosen will be referred to as the 'Off Road Vehicle' and is shown in Figure 8.1 (overleaf).

Figure 8.1. The Lego model of the 'Off Road Vehicle' used for the experiment

8.2.1. Pilot Studies

A number of pilot studies were carried out involving interviews and direct observation in order to examine how people learn how to assemble the model. Participants of the pilot studies were people who were within a similar working environment, generally other researchers. Three groups of four people were studied, all were told that they would have to make the model three times. The first time they made it, one group had the instructions (referred to now as Instructions group), the second group received a demonstration of the task (referred to now as Demo group) and the third group just had the picture of the model (referred to now as Picture group). During the second and third time the groups had to make the model, they were given no instructions at all.

The resulting times, on the first assembly, revealed that the Instruction group consistently took less time than the other two groups. In fact, the Picture group took about three times as long as the first group and the Demo group took about twice as long. However the second assembly showed a reverse of these results. The Instruction group took a lot longer than both the Picture and Demo groups. Even the third time, the Instruction group were slightly slower.

From informal interviews and direct observation what appeared to be happening during the initial assembly was that the Instruction group were ‘blindly’ following the instructions without being aware of the potential mistakes which could be made in assembling the model. When it came to making it without the instructions, after the first error they were immediately confused. Eventually they would abandon trying to remember the instructions and instead attempted to build the model from their image of the final model. The Demo group were able to remember about half of what they saw but then could not remember it all so again when an error was made they became confused. However the Picture group were initially made to solve the problem themselves so they learnt as much from their mistakes as from knowing where the pieces went. So when they had to make the model a second time they had already encountered the pitfalls. Therefore a training method which allowed the participant to practice a ‘hands-on’ discovery learning approach to the task appeared beneficial.

Even though this was a relatively generic task, it could still be considered against the VE selection criteria discussed in chapter six as follows:

- it is important for effective learning to experience the task in 3D through more than one sense (visual, aural, tactile)
- it is important for effective learning to have some ‘hands-on’ interaction with the real equipment or process
- it is important for effective learning to allow the trainee to practice a number of times on the real equipment or process
- it is important for effective learning to allow the trainee to set their own pace
- it is necessary for understanding the task to use a number of different view points
- current methods of training are inadequate in some way

The task involves recognition of parts, basic fine motor skills, some spatial awareness of where parts are and learning by rote a pre-defined sequence. These skills are learnt more effectively through experience of the task or at least a realistic representation of the task which allows the trainees to perform all the same actions as required in the real world.

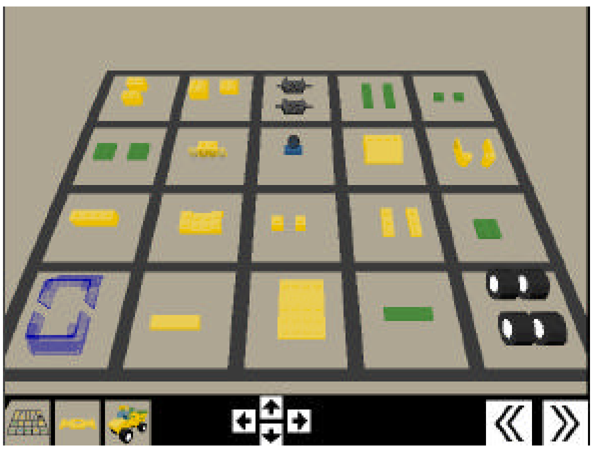
Also, while during the pilot studies the visual instructions were effective in showing how the task should be performed, some instructions could be misinterpreted and participants felt that they would have liked to see some parts from other viewpoints. Therefore the current method of training can be improved.

8.3 Specification Stage

8.3.1 Training Needs Analysis

As outlined in chapter six, before the development process of the VE begins some training needs analysis is required therefore task and person analyses were performed. In order that all the participants could learn the same procedure the model was laid out on a parts grid in a five by four formation, as shown in Figure 8.2. This is akin to tasks in some manufacturing processes where people are required to assemble components from a set of parts.

Figure 8.2. Parts grid of the Off Road vehicle



- *Task analysis*

To analyze the task each part was given a number and name as listed in Table 8.1.

Table 8.1: Parts list of the Off Road Vehicle

No.	Name	Qty	No.	Name	Qty

0.	Chassis	1	11.	yellow 4-block	1
1.	yellow 2-block	2	12.	rear wheel arch	1
2.	reverse slope	2	13.	yellow 1-block	2
3.	wheel axle	2	14.	yellow 3-flat	2
4.	green 4-flat long	2	15.	green 4-flat	1
5.	green 1-flat	2	16.	windscreens	2
6.	green 4-flat square	2	17.	yellow smooth flat	1
7.	Light mounting	1	18.	cantilever roof	1
8.	Steering mount	1	19.	green smooth flat	1
9.	Front wheel arch	1	20.	wheels	4
10.	Doors	2			

The steps involved in the task were then described as follows:

1. pick up part (0) chassis
2. turn it over
3. pick up one of part (1) yellow 2-block
4. turn it over
5. attach it to the right hand side of the base plate on the chassis
6. pick up another part (1) yellow 2-block
7. attach it to the left hand side of the base plate on the chassis
8. pick up one of part (2) reverse slope
9. attach it next to part (1) on the right hand side with the square edges flush to each other and the slope facing outwards
10. pick up another part (2) reverse slope
11. attach it next to part (1) on the left hand side with the square edges flush to each other and the slope facing outwards
12. pick up one of part (3) wheel axle
13. turn it over
14. attach it to parts (1) and (2) on the right
15. pick up another wheel axle
16. turn it over
17. attach it to parts (1) and (2) on the left
18. turn the whole object over and place it down so that it resembles a horizontal rectangle
19. pick up one of part (4) green 4-flat
20. place it horizontally on the upper part of the central square

21. pick up another part (4) green 4-flat
22. place it horizontally on the lower part of the central square
23. pick up one of part (5) green 1-flat
24. place it on one of the top studs at the end of the chassis
25. pick up another part (5) green 1-flat
26. place it at the same end of the chassis as before but on the opposite side
27. pick up part (6) green 4-flat square
28. place it next to part (5) so that their edges are touching
29. pick up another part (6) green 4-flat square
30. place it on the other side of the central square
31. pick up one of part (7) light mounting
32. place it between parts (5) in the remaining studs at the end of the chassis
33. pick up part (8) steering mounting
34. place it next to part (6) in the central square
35. pick up part (9) front wheel arch
36. place it on top of parts (5), (6) and (7) and next to part (8)
37. pick up one of part (10) door
38. place it on one side of part (8)
39. pick up another part (10) door
40. place it on the other side of part (8)
41. pick up part (11) yellow 4-block
42. attach it to the back of the chassis on the free studs
43. pick up part (12) rear wheel arch
44. place it in front of part (11)
45. pick up one of part (13) yellow 1-block
46. place it at the end of part (10)
47. pick up another part (13) yellow 1-block
48. place it at the end of the other part (10)
49. pick up one of part (14) yellow 3-flat
50. place it on top of part (12) and (13) on one side
51. pick up another part (14) yellow 3-flat
52. place it on top of part (12) and (13) on other side
53. pick up part (15) green 4-square flat
54. attach it to part (11) in the middle free studs so that it overhangs to the back of the vehicle

55. pick up one of part (16) windscreen
56. attach it to the front bonnet
57. pick up another part (16) windscreen
58. attach it to the rear of the vehicle
59. pick up part (17) yellow smooth flat
60. attach it to the rear part (16)
61. pick up part (18) cantilever roof
62. attach it to the front part (16)
63. pick up part (19) green smooth flat
64. attach it to the overhang part (15) at the rear of the vehicle
65. pick up one of part (20) wheels
66. attach it to the wheel axle part (3)
67. pick up another part (20)
68. attach it to another wheel axle part (3)
69. pick up another part (20)
70. attach it to another wheel axle part (3)
71. pick up another part (20)
72. attach it to another wheel axle part (3)
73. place the vehicle down into the box

As can be seen immediately, written instructions of the task are not only difficult to write but are also difficult to understand without some visual representation. Performing the task was simple enough to follow but could only be remembered after a few practice sessions. Also any errors had a 'knock-on' effect so that as soon as one step went wrong it had to be corrected, or subsequent steps would also be wrong. Therefore, what was required of the VET was to allow the trainees to practice the assembly at their own pace and in their own way.

- *Person analysis*

The skills required for this task were basic recognition of parts, procedural knowledge and fine motor skills.

8.3.2 Specification of the VE

A suggested format for a specification for the VE was outlined in chapter six and the sections which are relevant to this study (as outlined in chapter seven) are discussed as follows:

- *Description of the target training population*

The VET was aimed at people who were unaware of how to complete this task. (A more detailed description of the participants in the study is given later in section 8.5.2.3)

- *Overall aim of the project*

The overall aim of this project was the same as that in experiment one (chapter seven) - 'to design, build and evaluate a VET application' - but unlike the Netcard VET the application was not to be trained using a step-by-step approach but rather by providing options for different ways of learning, for the trainee to choose the one most suitable to them.

- *Objectives of the VET application*

Therefore the objectives were also similar to those in experiment one but slightly more focused as follows:

1) To highlight *specific* features of desktop VR/VEs that may be beneficial to training applications;
2) To train non-experts with the VET application to follow a procedure to assemble a toy car; and
3) To compare the value of VET over other methods of training.

- *Choice of VR system*

The system chosen was desktop VR consisting of a Pentium 133 PC, a 14-inch monitor display, keyboard and mouse with Superscape VRT 4.00 software.

- *Choice of VE Developer*

The VE again was developed 'in-house' by VIRART because of their experience with developing desktop VR applications with Superscape software.

- *How the training is to be evaluated*

As in experiment one (chapter seven) three groups would be used - a control group to show the value of the application against no training and a group trained by another method to compare some of the specific features of VET.

- *Time-scale of the project*

Given the simple nature of the task, two months were allowed for designing, building and testing the VET before carrying out experimental trials.

- *The real environment*

As this application was developed for research the real environment was an experimental laboratory.

- *Content of Virtual Environment*

The basic requirements of the VE are shown in Table 8.2.

Table 8.2. Basic content of virtual environment

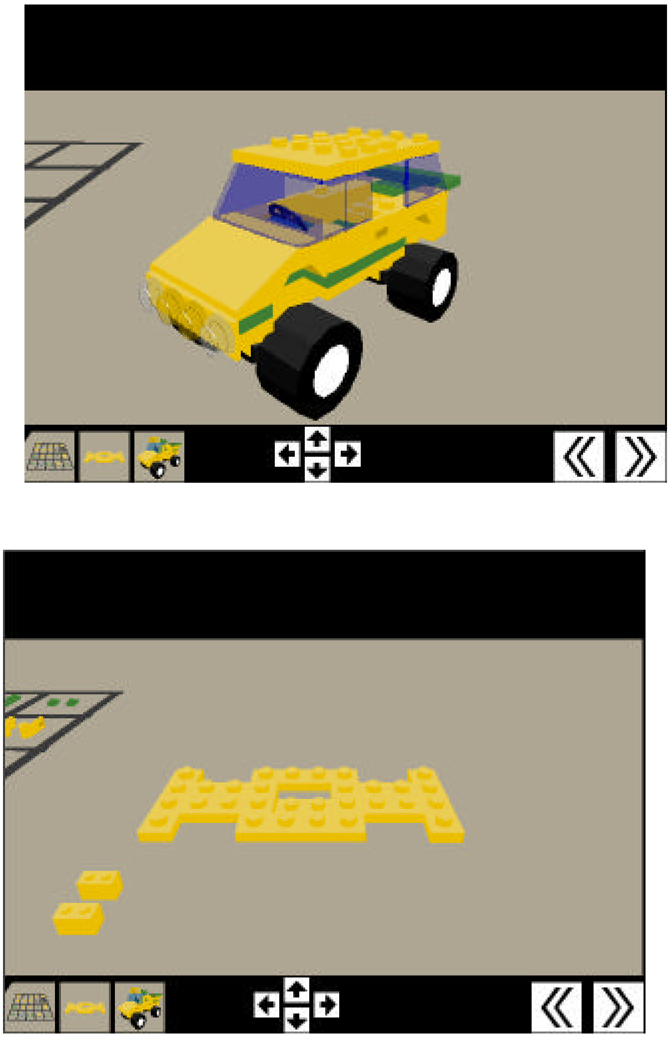
Features	**Required in VE**
• Objects	- a table - the finished model - the parts grid with each part in place - a construction area
• Functionality	- table: no functionality - the finished model: to be picked up and rotated so that all views can be seen. - the parts grid with each part: the parts to be picked up in turn and be attached to each other - construction area: the user should be able to pick up the incomplete model and rotate it so that all views can be seen
• User interactions	“pick up” a part “attach it” to another part

	"rotate" the objects to view from every angle
• Special requirements	- the trainee must be able to learn the task in the way they would like to through discovery learning.

8.4 Building the Application

The building stage involved collaboration between the VE developer and this researcher (as the proposer of the application) through numerous review sessions. These reviews concerned how to design the application given the specification outlined above (see also Eastgate et al, 1997b). The application will be referred to as the Lego VET. The layout of the screen is shown in Figure 8.3.

Figure 8.3. The layout of the Lego VET



The screen is divided into two - the main part at the top is the VE and the bottom half has the interaction devices. Firstly there were constraints of the system which had to be considered. Given the problems with the spacemouse encountered by many of the participants in the first experiment (see chapter seven) it was decided to use an on-screen device to allow users to rotate the objects. This device consisted of simple arrows which allowed the user to rotate the object along the x and y-axis. This device is located in the centre of the bottom layout on Figure 8.3.

Also, the computer monitor only had a field-of-view of 14 inches across, which was insufficient for the user to be able to see all the elements of the task laid out on the table. So the user in the VE was fixed to a seated viewpoint (as they would be seated in the real world) and in order to virtually rotate their heads from side-to-side to see the parts on the table, three viewpoint buttons were provided. These are located on the left of the bottom layout on Figure 8.3. The viewpoints consisted of the parts grid, the construction area and the completed model.

A second issue was to provide the trainee with different ways to learn how to perform the task. Written instructions were not included as it was felt that they were difficult to understand and therefore not useful (as shown in section 8.3.1). However problem solving/discovery learning and automatic demonstration were included. The real task required the operator to follow a pre-defined sequence so the objects in the parts grid could not be selected randomly. Instead, the user selected each one in turn and it automatically went to the construction area. Then, if they felt that they knew where the part should go or they would like to guess, they just clicked on the appropriate stud on the object already in the construction area. If they were correct the new part automatically attached itself. If they were incorrect, there was no response. In order to help the user the completed model is provided as a guide. Using the on-screen rotation device the user could examine the model more closely from all angles. Alternatively, if they really do not know where the part goes or they would like to see a particular action again, there are 'forward' and 'backward' buttons located at the right of the bottom layout on Figure 8.3. These take a step forward in the construction or a step backward. The trainee is then also able to learn how to build the model by seeing a full demonstration of how it pieces together, as well as being able to practice construction themselves.

8.5 Evaluating the Application

The same methods of evaluation used for experiment one (chapter seven) were used in experiment two. Formative evaluation was performed to examine the design of the application and then summative evaluation to examine the outcomes.

8.5.1 Formative Evaluation

The layers of formative evaluation (see chapter five) used were self-evaluation, expert review and non-expert review and a small group or laboratory experiment as follows:

8.5.1.1 Self-evaluation

Between this researcher and the VE developer, the Lego VET was examined for its design and content in view of the specification outlined before until we were satisfied with its development.

8.5.1.2 Expert and Non-expert review sessions

Five participants from the initial pilot studies were asked to assess the application for design and ease of use. These participants knew how to carry out the task in the real world therefore they were used as 'experts' of the task. It took approximately five minutes for them to familiarize themselves with the components of the application and after then they had little difficulties in performing the task in the VE. Feedback was also provided by a couple of members of the CREDIT (Centre for Research in Education and Training) group from the Department of Psychology at the University of Nottingham and three other novices to the task. They assessed the application and then performed the task in the real world. Generally the response was positive and relevant modifications were made.

8.5.1.3 Small group or laboratory experiment

Evaluation of the design was also carried out as part of the experiment outlined next.

8.5.2 Summative Evaluation

As the application was developed for research the only levels of the four levels of evaluation outcomes (Kirkpatrick, 1991 - see chapter five) that could be examined were Level 1: Reactions and Level 2: Learning outcomes. However unlike in experiment one (chapter seven) a cost benefits and cost-effectiveness analyses will not be performed as they will more or less be the same as the costing discussed at the end of chapter seven. Information for the levels was gathered through a controlled experiment as follows:

8.5.2.1 Research Objectives

The objectives of the VET application (stated in the specification 8.3.2) were as follows:

1) To highlight *specific* features of desktop VR/VEs that may be beneficial to training applications;
2) To train non-experts with the VET application to follow a procedure to assemble a toy car; and
3) To compare the value of VET over other methods of training.

In order to fulfil the first two objectives the application would need to show some transfer-of-training to the real task and the last objective would need a comparison of the performance and opinions of trainees of a VET and another method to be made. To do this, the application was measured against participants that received no training to show the benefits of VET and also measured against participants trained by an alternative method, in order to prove the value of using VET. Like the first experiment, this alternative method will be akin to a training video, as this is a common method for training such tasks. However this training video will be an automated version of the task being performed in the VE. This is to ensure consistency in the interface to the participants, so that they all see the same computer animated version of the model and the only difference between the training methods will be the interactivity permitted by VET. This alternative method will be referred to as the video demo. Given these objectives the hypotheses are as follows:

8.5.2.2 Experimental Hypotheses

H_1: Participants trained by the VET will perform the task faster than participants who have received no training.

H_2: Participants trained by the VET will perform the task faster than participants who have been trained by the video demo.

H_3: Participants trained by the VET will make less errors than participants who have received no training.

H_4: Participants trained by the VET will make less errors than participants who have been trained by the video demo.

H_5: There will be a difference in the opinions of the participants effected by the training method experienced

8.5.2.3 Participants

Thirty-six people participated in the experiment with an equal divide of male and female participants. Ages ranged from 22 - 44 years of age with an average of 28.6 years of age. They were either self-selected as respondents to posters placed around the University or recruited by direct contact and ranged from secretarial and library staff to research assistants and associates from a variety of disciplines including Manufacturing, Chemistry, Biology, Gynaecology, Medicine, etc. They were required for two sessions lasting about 60 minutes with five days between each session. They were paid £10 for their total participation. Selection was based on their experience of Lego and other model kits. It was found through initial pilot studies that those who had played with Lego recently (usually, if they had small children or looked after small children), appeared to build models without instructions quicker than those who had not used Lego for at least five years. This also eliminated the likelihood that they would have made the off road vehicle beforehand, as it is a relatively recent model. Therefore, participants were selected if they had ‘no’ experience of Lego or other model kits in the last five years and if they didn’t have young children who played with Lego.

8.5.2.4 Independent and dependent variables

The independent variable was again ‘training’ and this was measured at three levels: (i) training with the VET application; (ii) training with a video demo; and (iii) a control group which received no training. The dependent variables were time, errors and opinions.

8.5.2.5 Measures

- Time was measured by a digital stop watch.
- Errors in performance of the task in the real world was measured through direct observation and video analysis. An error was considered to be any piece of Lego that was in the wrong place.
- Opinions were measured through questionnaires consisting of open questions, multiple choice and rating scales at the beginning, during and at the end of the experiment.

Further information about the design of these measures and copies of each can be found in Appendix (IV).

8.5.2.6 Design and Procedure of Experiment

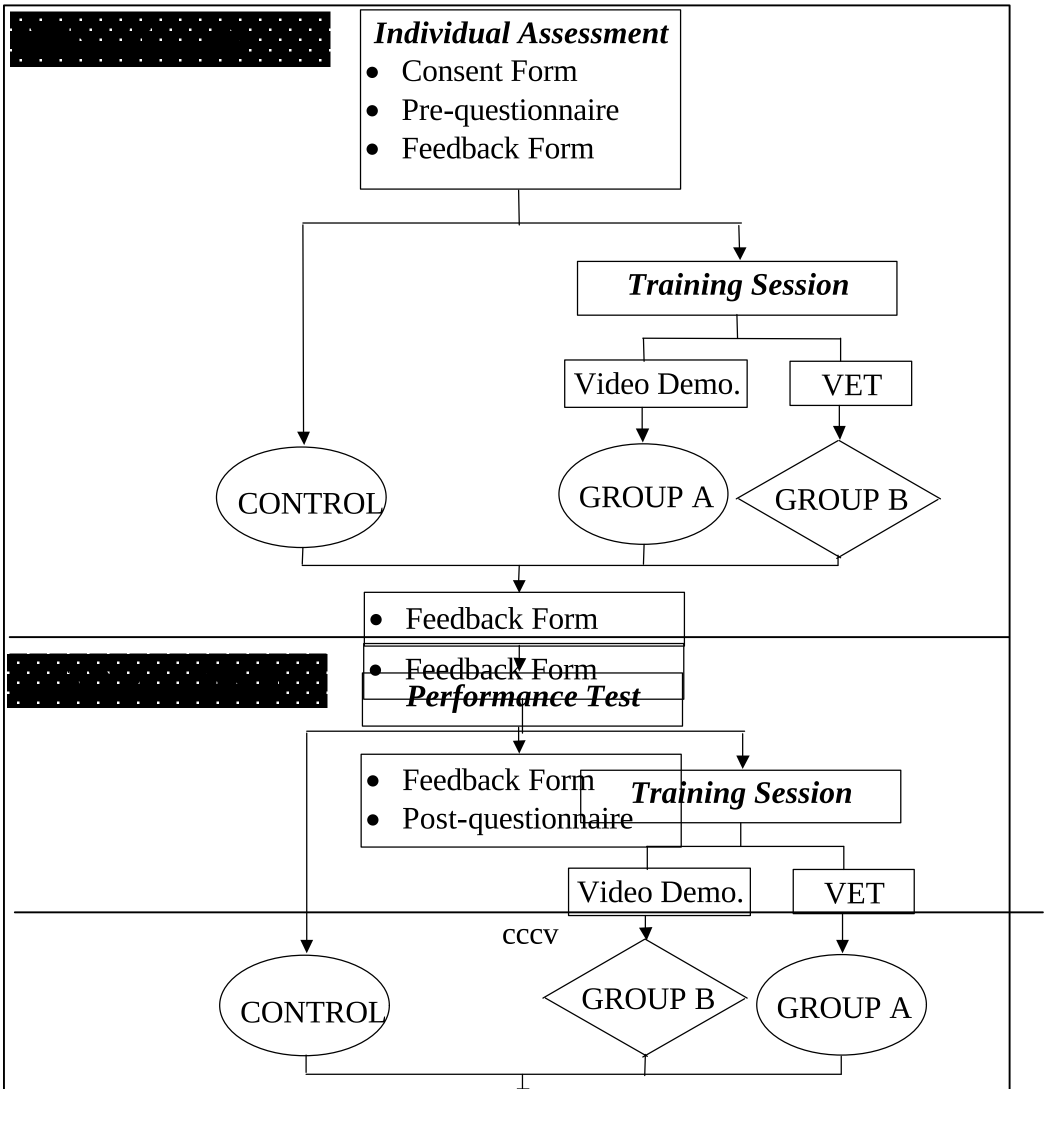
Both a within-subjects and between-subjects design was applied, in order to compare the results of the groups and to allow the participants to compare the training methods themselves. The procedure of the experiment is illustrated in Figure 8.4 (overleaf) and discussed as follows under the headings - individual assessment, training session and performance test.

(i) Individual Assessment. All the participants were required for two sessions. In the first session the participants completed a consent form providing them with information about the task and a pre-questionnaire which collected information about their experience and attitude towards computers. The participants were then randomly divided into three groups - a control group, group A and group B - although an equal divide of male and female was maintained in each group. The control group were told that they would have to assemble a toy car five times

without any instructions. However they would have five minutes of pre-exposure of what the car was to look like before the task was carried out (this is discussed below). Group A and Group B were told that they would be trained with two different computer methods to learn a basic procedure for assembling a toy car. Then after training they would be performing the task five times but no further instructions would be provided.

When the participants were given their relevant information they were asked to complete a feedback form which consisted of six semantic differentials on a scale to indicate how they felt about the task. This feedback form was again administered after training and then again after the performance test. Also after the test they completed a post-questionnaire and if they had been trained by the VET application they completed a further questionnaire on the features of the VET application.

Figure 8.4 Procedure of Experiment Two



In the second session, the participants completed the feedback form again before training, after training and after the performance trials. Then they completed a post-questionnaire and finally a questionnaire allowing them to compare the two tasks and two training methods. Further information about the design of these questionnaires and copies can be found in Appendix (IV).

(ii) Training stage. In the two sessions the Control group received no prior instruction before carrying out the task. However because during the performance trials the final model was not present for any guidance, some pre-exposure was required so that the participants had some idea of what they were to make with the parts. Therefore as the time limit for training was set at ten minutes and the other two groups only got to see the completed model in computer form beforehand, the Control group were also exposed to the completed model in computer form, in rotation (so that all angles could be seen) for ten minutes.

In the first session, Group A were trained to assembly the off road vehicle by the video demonstration and Group B were trained by the VET application. The video demonstration consisted of the participants being seated in front of the computer monitor and watching an automated sequence of the task for ten minutes. This enabled them to see the assembly sequence five times. Training was set at ten minutes so that every participant had the same exposure time to the completed model. It was also shown through the review sessions that ten minutes was an adequate time for the participants using the VET application to explore all the options offered to them and to perform the assembly sequence at least three times. VET involved an introduction to the layout and options on the screen, five minutes to familiarize themselves with the layout and then ten minutes to train themselves.

In the second session a comparable model had to be found as the participants had already been trained to make the off road vehicle. After testing a number of models, it was decided to create the second model by reassembling parts of the first model. This was called the 'monster truck'. As well as having similar parts it also consisted of the same number of steps as the first model. Group A were then trained by the VET application and Group B were trained by the video demonstration. In this way both groups could experience both training methods.

(iii) Performance Test. After training all the participants carried out the same performance test which involved actually completing the real task. The task was laid out exactly as they had seen in the training stage. The participants were told that they had to assemble the model five times as quickly and as accurately as possible following the procedure. Each part had to be picked up in turn and placed on the main chassis before moving on to the next part but if they realized later that the part was in the wrong place, they were allowed to take the model apart and move it. However they were not allowed to carry on to the next part until the previous part had been attached to the main object. After the first trial if errors had been made the participants was told so, and the experimenter went through the correct procedure once for them. The participant then did the second trial and if errors were made, the experimenter went through the correct procedure again. This continued until the five trials were completed. The number of trials was set at five as again from the pilot studies it was shown that it took at least five trials for all the participants to be able to complete the model without errors. Each trial was observed, video recorded and timed for later analysis.

8.5.2.7 Results

The results will be discussed using the three stages of data collection outlined in the procedure. Throughout the groups will be referred to as Control group, Group

A (participants which performed the video demo in the first session then the VET) and Group B (participants which performed the VET in the first session then the video demo). A more detailed breakdown of the results can be found in Appendix (IV) but is briefly outlined as follows:

(i) Individual assessment. A summary of the subject profile is shown in Table 8.3.

Table 8.3: Profile of subjects per group in experiment two

	Control	Group A	Group B	Total
Average age	28.8	28	26.3	28.6
Gender	male (6) female (6)	male (6) female (6)	male (6) female (6)	male (18/36) female (18/36)
Use of computers	everyday	everyday	everyday	everyday (31/36)
• Games	some	some	some	some (25/36)
• Wordprocessing	alot	alot	alot	alot (29/36)
• Spreadsheets	some	some	some	some (21/36)
• Design	none	none	none	none (24/36)
• Databases	some	some	some/alot	some (17/36)
• Internet	some	some	some/alot	some (17/36)
• Programming	none	none	none/some	none (21/36)
• Maintenance & Repair	none	none	none	none (31/36)
Virtual Reality	yes(12/12)	yes (12/12)	yes(12/12)	yes (36/36)
• newspapers/magazines	9	7	9	25/36
• TV, film, video	7	9	11	27/36
• observed use	4	6	6	16/36
• experienced use	4	5	6	15/36
• worn headset	4	6	5	15/36
• played game	1	1	2	4/36
• other application	2	3	4	9/36
Computer Attitude Scale	126.75	131.67	126.67	128.36
Experience of Lego	yes(10/12)	yes (11/12)	yes(12/12)	yes (33/36)
Last experience years	11 - 15	11 - 15	11 - 15	11 - 15 (14/36)
Rating of experience	little	some	some	some (14/36)

As Table 8.3 shows the average age of the participants was 28.6 years old and there was an equal divided of male and female participants in each group. The majority of the participants used computers everyday (31/36) and in terms of their type of experience, the groups were fairly evenly matched with the majority in each having ‘a lot’ of word processing experience (29/36), ‘some’ games (25/36) and spreadsheets (21/36) experience but no experience of design (24/36) or maintenance and repair (31/36). Group B though had slightly more experience of the internet and databases and some experience of programming than the other two groups. All the participants had heard of the term ‘virtual reality’ mostly from the television, video and films (27/36) and newspapers and magazines (25/36). About two fifths had actually watched someone using a VR system (16/36) and used a system themselves (15/36), as well as worn a headset (15/36). Their experience appears to be either of playing a VR game (4/36) or some other application other than games (9/36). Given this type of experience most of the participants had a good understanding of VR with the words ‘real or realistic’ (21/36) ‘computer generated’ (15/36), ‘3D’ (10/36), ‘simulation’ (10/36) and ‘interactive’ (6/10) being the most frequent words used in their definitions. Applying the Computer Attitude Scale (developed by Loyd and Gressard, 1995) on the whole all the participants had a high positive attitude towards computers with little differences between the sub-scores.

Only three participants had no experience of using Lego before. Two were in the control group and one was in the group A. The last time that the participants had used lego ranged from ‘less than 5 years ago’ (2/36) to ‘more than 20 years ago’(5/36). However more of the participants were between 11 - 15 years (14/36) then between 6 - 10 years ago (7/36) or 16 - 20 years ago (6/36). The types of experiences they had were generally making models with and without instructions when they were children themselves or helping children make models. The rating of experience went from ‘none’ (3/36) to ‘a lot’ (7/36). In the main more of the participants rated their experience as ‘some’ (14/36) then ‘very little’ (5/36) to ‘little’ (7/36). There were slight differences in the groups with the control group rating themselves with slightly less experience then groups A and B. Therefore as can be seen from Table 8.3 the groups were fairly well balanced in their experience of computers and Lego.

(ii) Training

In the comments made by each group about the training methods there were only slight differences based on which method they experienced first. Where these differences occur, they will be highlighted, otherwise discussion will proceed considering the information from both sessions as a whole.

Video demo training - As group A were trained by the video demo first, their initial impressions of the training were slightly more positive than group B's impressions of the video demo. On the whole group A's impressions of the video demo were fairly positive with the comment "*easy to understand*" used frequently, but a few found it initially too fast (3/12). All the participants (12/12) said that they had benefited from the training mainly because the repetition enabled them to remember the task and the skill was transferred to the real situation. The majority (11/12) said that it gave them all the information required to carry out the task, although one participant felt that it didn't show you that you had to think about which way the pieces had to be turned around to be attached. As for group B's impressions of the video demo these were mixed. Some found it particularly repetitive and boring especially as no interactivity was permitted. A few of the participants thought the viewpoints were awkward and another thought it was too fast. However one participant preferred the video demo and others felt it was useful. Three-quarters of the participants (9/12) felt that they benefited from the training because the repetition helped to remember the sequence of the task. However a quarter of the subjects (3/12) did not feel that they benefited because they felt that it was 'boring' and failed to hold their interest. The majority of the participants (11/12) felt that the training programme gave them the required information however one participant felt that you couldn't view the building process from all aspects. Regarding improvements, suggestions included: rotating each new part before it is attached, fewer repetitions, slower speed especially the first time, 'action music', varying views of the process and some interaction.

VET application - Impressions of the VET were generally positive with words like 'interesting' and 'fun' used frequently and comments made like *"I felt far more confident about building the model from scratch ... "* and *"better in many ways then the previous session because I could go at my own speed and concentrate on bits that I felt I may forget"*. One participant particularly liked the viewpoints and a couple said that it maintained their concentration. However one participant felt that it would have been nice to have been able to put the pieces in the wrong place because it would have been more realistic and another felt that the training was more complicated than the task. A couple of participants felt that a *"... mix between 1st and 2nd would have been easier for me"* and *" ... preferred the video it seemed as though you had more time, the computer based system was more interesting."* The majority of the participants however (22/24) felt that they benefited from the training mainly because it allowed them to go through difficult steps a number of times and reinforced the sequence in their minds and they were able to transfer the skill to the real situation. One participant felt that they had not benefited as they preferred *"... interaction with people rather than computers"* and another felt that they spent a lot of time learning how to use the computer rather than learning how to perform the task. (This participant had had no previous experience of computers). The majority of the participants (22/24) felt that the training programme provided them with all the required information though two participants felt that it was difficult to remember all the information in such a short space of time even though the time was the same for both training methods.

Regarding improvements some participants felt that it would have been better if they could see all the objects on the table at the same time. However, as explained in section 8.4, because the system used a standard-sized 14 inch monitor it was difficult to view all the objects at once because this would involve being at a high viewpoint. This would make the Lego pieces very small and difficult to manipulate and the user would have to keep ‘zooming’ in closer to see the objects anyway. Therefore the three viewpoints were provided as a solution and were akin to turning one’s head to look at the different objects. Another alternative would be to provide a different display system with a wider field-of-view.

Another suggestion was to provide a further response if their actions were correct or incorrect. The current response was either the part attached itself to the rest of the blocks or did nothing, thus reinforcing the correct procedure. One of the initial versions of the application included the use of error prompts. If the user guessed incorrectly a prompt would appear and they would then have to click confirming it had been acknowledged before being allowed to proceed. During the review sessions the participants found this particularly distracting as they were constantly having to stop to remove the prompts and this interrupted the learning process. Equally an error noise was found not only distracting but also embarrassing as participants felt that other people were aware that they were making mistakes. As these responses appeared to hinder the learning process it was decided to remove them so that the participant could freely learn the process in their own way without any interference. However some people do prefer some form of response therefore a future modification would be to provide the user with the option of turning this response on and off. Other suggestions included the ability to make mistakes and to be able to drag and rotate the part before attaching it to the rest. These will also be considered as future modifications.

Considering the rating scales used in the VR/VE assessment tool discussed in Appendix (II) the results are shown in Tables 8.4 and 8.5 (overleaf).

Generally the participants from both groups were satisfied with the desktop VR system and agreed with each of its components – the monitor, the mouse and the on-screen rotation device. They also disagreed that a headset would enhance the system. As for the VE, again they were generally satisfied with all of its parts and agreed that each aspect was fine.

Comparison of Video demo and VET - A list of eight descriptive words rated using a Likert scale in order to obtain parametric data were administered to compare what the participants thought of the applications. The results are shown in Tables 8.6 and 8.7. As these tables show, there were no significant differences within-groups in their opinions of both training methods. Generally they agreed that both methods were relevant, helpful, effective and useful; and they disagreed that they were inappropriate, too slow, unhelpful or confusing.

Table 8.6: Comparison of the opinions of Group A of the Video demo and VET

GROUP A			
Factor	**Video demo.**	**VET**	**Significance**
Relevant	Strongly agree	Strongly agree	t = -.761; df = 22; p>0.05
Helpful	Strongly agree	Strongly agree	t = -.266; df = 22; p>0.05
Inappropriate	Strongly disagree	Strongly disagree	t = -1.06; df = 22; p>0.05
Too slow	Slightly disagree	Slightly disagree	t = -.688; df = 22; p>0.05
Effective	Strongly agree	Strongly agree	t = -1.119; df = 22; p>0.05
Unhelpful	Strongly disagree	Strongly disagree	t = -.673; df = 22; p>0.05
Confusing	Strongly disagree	Strongly disagree	t = -1.121; df = 22; p>0.05
Useful	Strongly agree	Strongly agree	t = -.321; df = 22; p>0.05

Table 8.7: Comparison of the opinions of Group B of the Video demo and VET

GROUP B			
Factor	**Video demo.**	**VET**	**Significance**

Relevant	Agree	Agree	t = .340; df = 22; p>0.05
Helpful	Agree	Agree	t = -.340; df = 22; p>0.05
Inappropriate	Disagree	Disagree	t = -.462; df = 22; p>0.05
Too slow	Neutral	Slightly disagree	t = -.968; df = 22; p>0.05
Effective	Slightly agree	Agree	t = .934; df = 22; p>0.05
Unhelpful	Disagree	Disagree	t = -.579; df = 22; p>0.05
Confusing	Disagree	Disagree	t = .616; df = 22; p>0.05
Useful	Agree	Agree	t = .842; df = 22; p>0.05

However there were some significant differences between-groups in terms of strength of feeling (although the opinions were in the same direction). For the video demo., group A found the method significantly more effective than group B (t = 3.957; df = 22; p<0.05) and for the VET, group A found the method significantly less confusing than group B. This may be due to order effects. As group A performed the video demo. training first their opinion was not effected by a comparison of another method then when it came to performing the VET they had a better idea of what was important to learn and therefore were able to use the VET more readily. Whereas group B performed the VET first and were faced with having to understand the system as well as learn the task.

Furthermore the feedback forms that were issued before training, after training and after the performance test were analysed through one-way ANOVAs (including the data from the control group) to produce the results shown in Table 8.8 (overleaf). The purpose of the forms was to highlight if there were any change in opinions of the participants which could be attributed solely to training and not effected by the performance trials. As Table 8.8 shows, group A changed their opinions throughout the session on whether the task would be easy and whether they felt confident and unconcerned about it. Further post hoc comparison's using Tukey's HSD test revealed that for the Video demo. the differences were between forms 1 and 2 in terms of easiness of the task (p = .012); and between forms 1 and 3 in terms of confidence (p < 0.001) and concern (p = 0.42). This appears to imply that after training with the video demo., group A felt that the task would be easy however they only felt confident and unconcerned about the task after the performance test. As for the VET, the differences were between forms 1 and 3 for both ease (*p* = .001) and confidence (*p* = .000) implying that an effect only occurred after the performance trials.

Table 8.8: A comparison of the feedback forms of the groups for each method of training

GROUP A		
Method	**Factor**	**One-way ANOVA**
VIDEO DEMO.	Interesting	F = 2.210; df = 2,33; p>0.05
	Easy	**F = 4.694; df = 2,33; p<0.05***
	Unstressful	F = 1.058; df = 2,33; p>0.05
	Enjoyable	F = 1.967; df = 2,33; p>0.05
	Confident	**F = 9.582; df = 2,33; p<0.05***
	Unconcerned	**F = 3.523; df = 2,33; p<0.05***
VET	Interesting	F = .203; df = 2,33; p>0.05
	Easy	**F = 8.489; df = 2,33; p<0.05***
	Unstressful	F = 2.567; df = 2,33; p>0.05
	Enjoyable	F = .433; df = 2,33; p>0.05
	Confident	**F = 9.018; df = 2,33; p<0.05***
	Unconcerned	F = 2.368; df = 2,33; p>0.05
GROUP B		
Method	**Factor**	**One-way ANOVA**
VIDEO DEMO.	Interesting	F = .538; df = 2,33; p>0.05
	Easy	F = 2.343; df = 2,33; p>0.05
	Unstressful	F = .219; df = 2,33; p>0.05
	Enjoyable	F = .071; df = 2,33; p>0.05
	Confident	F = 2.870; df = 2,33; p>0.05
	Unconcerned	**F = 3.770; df = 2,33; p<0.05***
VET	Interesting	F = .816; df = 2,33; p>0.05
	Easy	**F = 6.836; df = 2,33; p<0.05***
	Unstressful	**F = 3.925; df = 2,33; p<0.05***
	Enjoyable	F = 1.467; df = 2,33; p>0.05
	Confident	**F = 6.712; df = 2,33; p<0.05***

	Unconcerned	*F* = 1.345; *df* = 2,33; *p*>0.05

Where * indicates a significant result

As for group B which performed the VET first, they had a change of opinions about ease, stressfulness and confidence. Tukey's HSD test revealed that there were significant differences between forms 1 and 2 and forms 1 and 3 in terms of easiness (*p* = .012 and *p* = .006; respectively); between forms 1 and 2 in terms of stressfulness (*p* = .041); and forms 1 and 3 in terms of confidence (*p* = .002). Therefore the VET appeared to have had an effect on the opinions of the participants to the extent that they felt the task would be easier and less stressful than they first imagined. As for when they performed the video training, the only difference was in concern between forms 1 and 3 (*p* = .032) i.e. it was only after the performance trials that they were unconcerned about the task.

Regarding descriptions of the video demo., mainly quite negative words like 'boring' (7/24) 'non-interactive' (6/24) 'just observing/watching' (10/24) were used. However descriptions of the VET were fairly positive with words used like 'interactive' (12/24) 'self-pacing' (6/24), 'in control' (4/24) and 'interesting' (4/24). As such all but one participant (23/24) said that they preferred the VET. The following comments summarise the general opinions:

" *VET allowed interaction and gave me a better 'feel' for model and the steps required to build it. It seemed a more natural learning process as you got to 'play' with the model as you went along*."

"*Attention span was longer with VET as you had to be more involved."*

"More interesting. I felt included in the training."

In terms of similarities and differences, the main ones suggested by the participants themselves are shown in Table 8.7.

Table 8.9: Similarities and Differences between the Video Demo and VET.

Similarities	**Differences**
• representation of the models • set-up of the task	• speed of training time • no control vs. control of pace

• procedure of the task • effectiveness of training	• passive vs. active participation • observation vs. 'hands-on' interaction • more to learn in the VET • more viewing angles in VET

The participants felt that both methods showed a good representation of the models and the intended task and procedure, as well as both being effective for training. However they felt that the training time went more quickly when using the VET than when watching the video demo. Also they had no control of their pace with the Video demo but were able to stop and start the procedure in the VET. They felt that the video demo was just observation and therefore passive, whereas the VET was like 'hands-on' interaction and therefore required active participation. However this meant that there was more to learn in the VET as they had to learn how to use the system as well as learn the task. Finally the VET provided more viewing angles than the video demo.

The participants were also asked what types of training applications they think would benefit from the use of VEs. These are listed as follows:

- Medical procedures (7) e.g. surgical techniques, analysis
- High-risk, dangerous, inaccessible environments (8/24) e.g. combat, fire-fighting, mountain rescue
- Basic manual jobs (6/24)
- Assembly tasks (5/24) e.g. D.I.Y, car maintenance, weapons assembly
- General training of tasks (5/24) e.g. procedural, computer tasks, everyday skills, repetitive, visualization
- Construction (5/24) e.g. planning, design of buildings
- Engineering (3/24) e.g. mechanical, motor vehicle
- Maintenance tasks (3/24)
- Learning to drive or fly (3/24)
- Education (3/24)

- Others (2/24) e.g. instructions for using domestic appliances, product design and testing

(iii) Performance test

The results of the performance times of each group per session are shown in Figures 8.5 and 8.6 respectively.

Figure 8.5: Boxplot of performance times of session one per trial per group

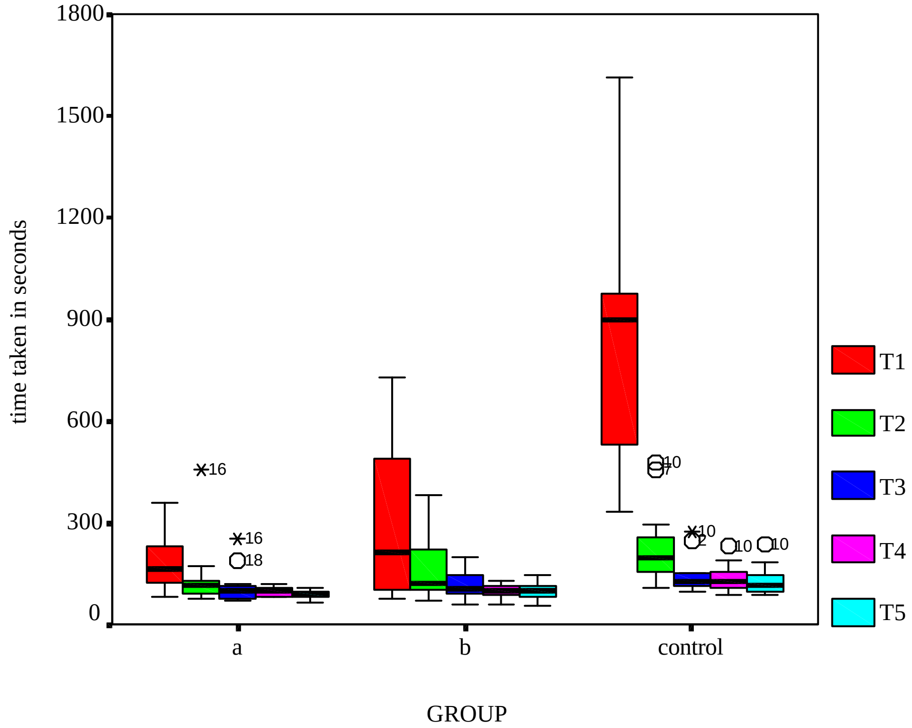
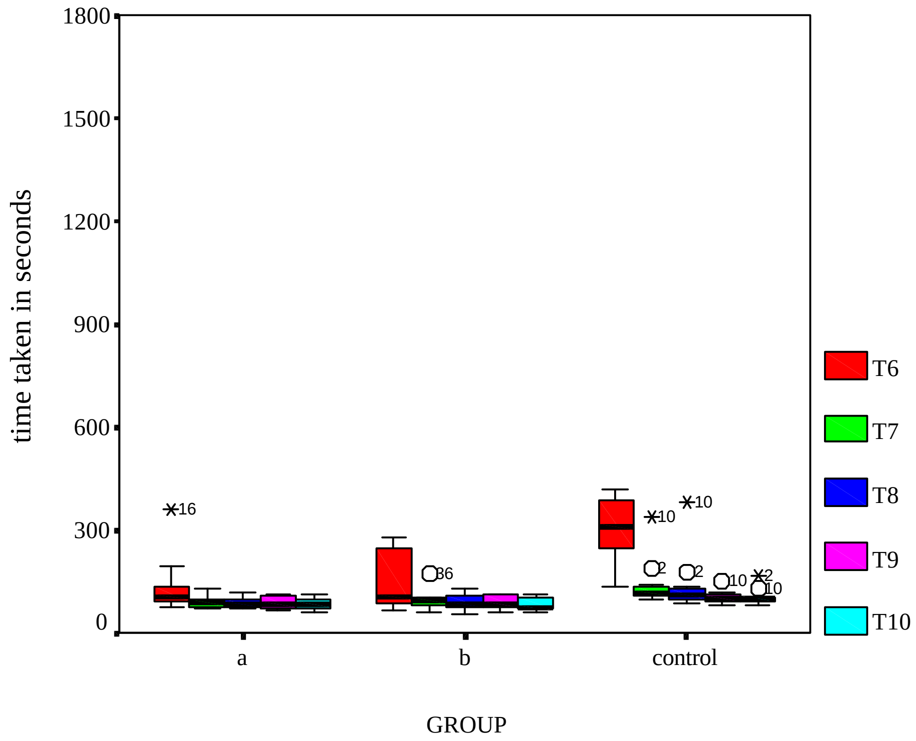


Figure 8.6: Boxplot of performance times of session two per trial per group



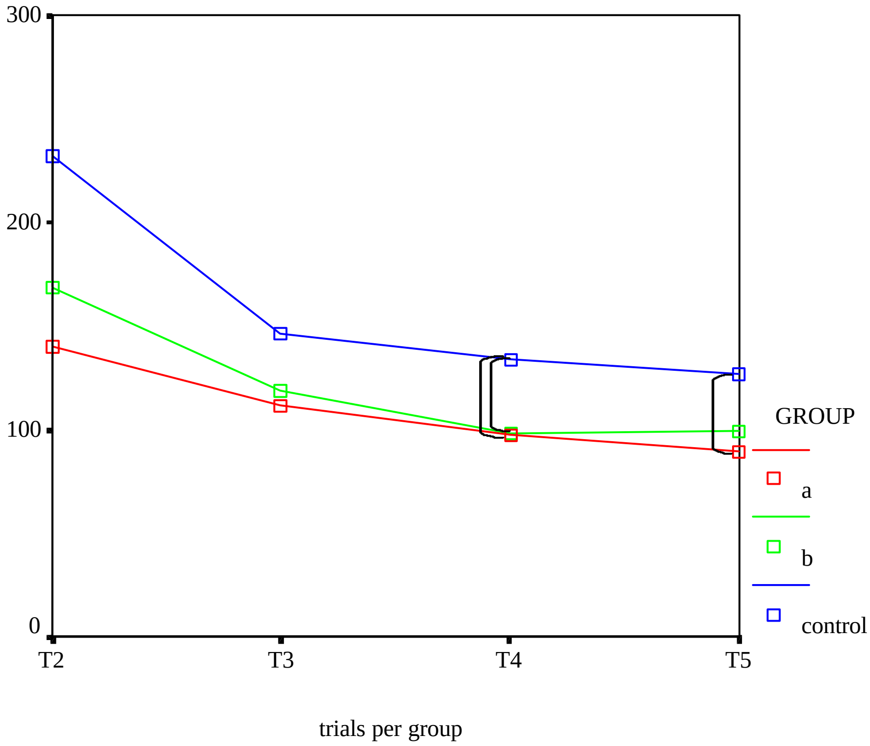
As can be seen, for all three groups the initial time in the second session is significantly lower than in the first session. This implies that the results of the second session are affected by the first, therefore the time taken within subjects per group cannot be statistically compared. However what can be noted is that during the first session there is large variability between the groups including a few outliers represented by 'o's and extremes represented by 'x's. This variability is gradually decreased by the end of the fifth trial. By the second session this variability is greatly reduced. Although the control group still take a longer time to complete the second task in the first trial.

In session one as predicted by the first experimental hypothesis – H_1: Participants trained by the VET will perform the task faster than participants who have received no training – both training groups outperform the control group especially during the first trial. This was confirmed by performing a repeated measures ANOVA firstly with trials as the with-in subjects factor to examine effects by trials, then with trials as the within-subjects factor and training as the between-subjects factor to examine effects by group per trial.

A significant difference was found for trial ($p = 0.003$). In particular between trial 1 and 3 ($p = 0.018$); trial 1 and 4 ($p = 0.036$) and trial 1 and 5 ($p = 0.017$). However the first trial can be considered different from the subsequent trials as it only has the effects of training and not added 'practice' effects. Applying tests to just the subsequent trials shows no significant difference ($p = 0.253$), therefore after the first trial there is no subsequent effects of trial on the results.

A significant difference however was found between the groups by trial ($p = 0.000$). Post-hoc comparisons using Tukey's HSD test revealed differences between the control and group A in trial 1 ($p < 0.001$); trial 4 ($p = 0.010$); and trial 5 ($p = 0.015$). Significance was also found between control and group B in trial 1 ($p < 0.001$) and trial 4 ($p = 0.015$). This is illustrated in Figure 8.7. Trial 1 has been excluded, as Figure 8.5 already clearly shows that it is significantly different by trial and group.

Figure 8.7: Average time results of the last four trials in session one



It appears that even though the control group continue to make an improvement on the time taken to complete the task, group A and B appear to make a better improvement than the control group in trial 4 and group A continue this improvement in trial 5. This could be accounted for by training providing a better retention of information which is enhanced by practice. Alternatively it could be the outliers making the difference. The idea that the training methods may have provided better retention is also shown in the results of the second session (illustrated in Figure 8.6). The control group having received no training again, still have slower initial times than the training groups. This appears to show that even though all three groups have obvious 'learning' effects from session one, training is still beneficial certainly in the first trial.

Considering the training groups, there were no significant differences between the performance times of group A and B in any trial or in any session. Therefore the second hypothesis - H_2: Participants trained by the VET will perform the task faster than participants who have been trained by the video demo - cannot be accepted. In fact it appears that in terms of performance times the VET is just as good as the video demo and not better.

The results of errors per group were considered separately from performance times. The errors per group per session are shown in Figures 8.8 and 8.9 respectively (overleaf). As can be clearly seen after the first two trials in session one, there are no errors made by any of the groups. Applying a repeated measures ANOVA confirmed significant differences for trial ($p < 0.001$) and between the groups by trial ($p < 0.001$). Further tests revealed significant differences between the control and group A in trial 1 ($p = 0.001$) and in trial 2 ($p = 0.047$) and between control and group B in trial 1 ($p = 0.042$).

Figure 8.8: Boxplot of errors in session one per trial per group

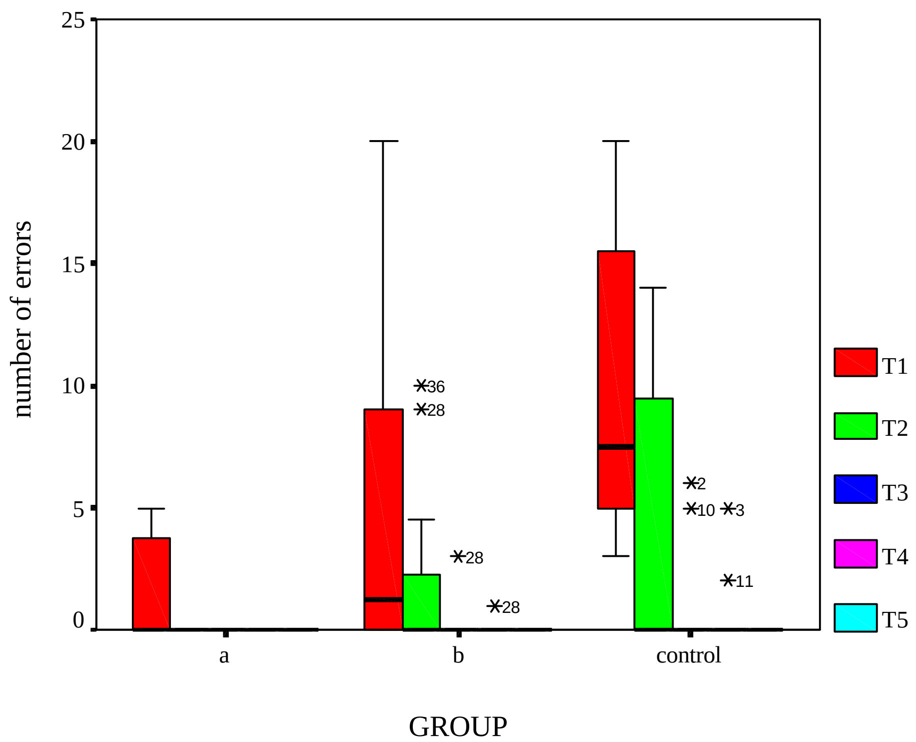
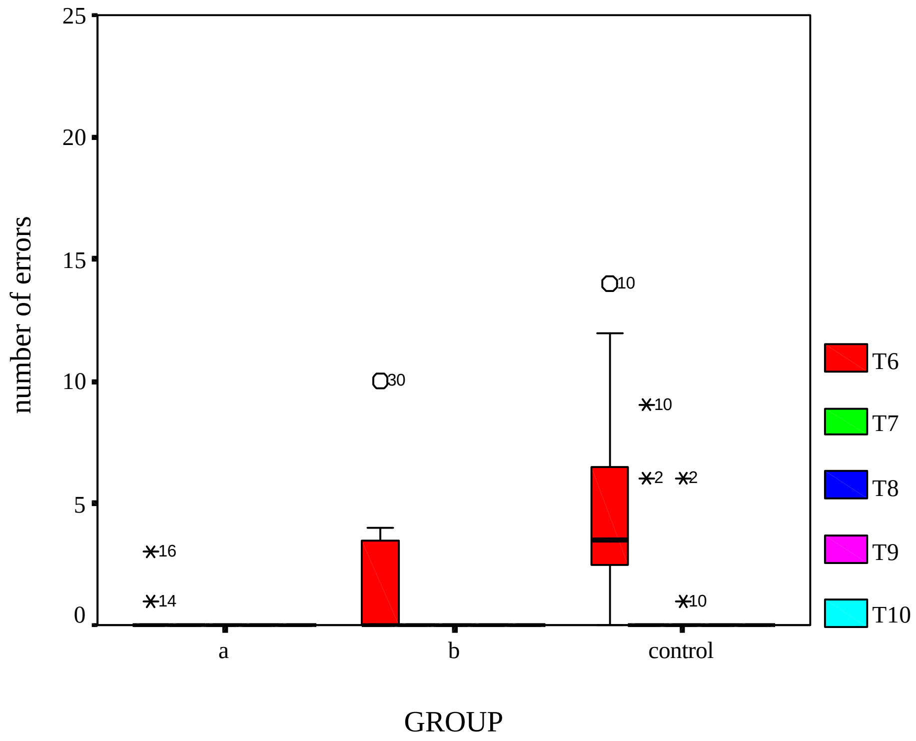


Figure 8.9: Boxplot of errors in session two per trial per group



Therefore the third hypothesis - *H_3:* Participants trained by the VET will make less errors than participants who have received no training - can be accepted for at least the first trial.

There were no differences between group A and group B in either trial in the first session therefore the fourth hypothesis - *H_4:* Participants trained by the VET will make less errors than participants who have been trained by the video demo - is rejected. Again the VET proves to be just as effective as the video demo in reducing errors, certainly in the first session.

As for the second session, the control and group B are still making mistakes in trial 1 while group A appear to be able to complete the task after training successfully. Group A in session one had been initially trained by the video demo. Then in session two they were presented with the VET. As they were aware of how the session would proceed from their experience of session one, they appeared to know the kinds of problems with the task that they were likely to encounter. Therefore as the VET allowed them to explore the task in detail they were able to explore the steps of the task that were particularly difficult. They were then more successful when performing the task. Group B however were initially trained by the VET. With little knowledge of the task and of using VET it was difficult for the group to obtain any real benefits from training. This can be clearly seen in Figure 8.8. Even by the second trial with 'practice' effects the group have difficulty. Then in session two, group B are trained by the video demo. Even though they have the experience of the first session they still make mistakes in the initial trial. This could either be because they did not benefit from the video demo as they were unable to interact with the model to understand the difficult steps, or it could be due to the varying abilities of participants within the two groups. These observations would suggest that a combination of conventional training and VET would greatly benefit a trainee.

8.6 Discussion

As in the previous chapter (chapter seven), this chapter examines the development process of a VET from proposal to evaluation but in more detailed. The objectives leading on from experiment one were as follows:

(1) To highlight *specific* features of desktop VR/VEs which may be beneficial to training

(2) To train non-experts with the VET application to follow a procedure to assemble a toy car

(3) To *directly* compare the value of VET over another training method.

Again before these are discussed the results will be considered under the evaluation levels of ‘reaction’ and ‘learning outcomes’ as follows:

Level 1: Reactions Level

The results showed that the video demo. was considered effective but not very interesting and limiting in the sense that only one view of the environment and its objects is given. In comparison the comments about the VET were generally positive. The participants were satisfied with the system and the features of the VE. Generally they found the VET interesting and highly interactive. The feedback forms highlighted a change in opinions after receiving training. After both the VET and video demo. the participants felt that the task would be easier. However given the opportunity to directly compare VET with a traditional approach to training – video demonstration – the majority of the participants (23/24) preferred the VET. This is because it allowed the trainee to have more control over how they wanted to learn.

Level 2: Learning Outcomes

The results illustrated in Figures 8.5 and 8.6 show that both training methods proved to be effective in reducing the learning time it takes for the trainee to 'master' the task. Both groups had faster performance times which were particularly significant in the first trial but they also continued to make a better improvement in subsequent trials (4 and 5). This though may be due to individual differences between the groups. Therefore it was possible to accept the first hypothesis - but reject the second - . The VET proved to be just as effective as the video demo.

In terms of errors, in the first session as illustrated by Figure 8.8, again both training groups prove to be significantly different to the control. Although all groups after the second trial do not make any errors. While group B which was trained by the VET may not be statistically different to group A, as Figure 8.8. shows some of the participants are still making errors by trial 2. This may be due to the fact that VET is a new training medium that they have generally not had previous experience of. Therefore when faced with a new technology some of the time effort must go towards understanding how to use it effectively. As there was a time limit placed on the experiment, some of the participants may have required longer to get to know the training before using it to learn how to do the task. Then when it came to the task there were steps that had not been retained and by the second trial these lessons were still being reinforced. In summary, some of the participants who may not have benefited from the VET would have begun the task as if they were the control group. Alternatively it could be due to individual differences as possibly illustrated in Figure 8.9. Some of the participants in group B, even though trained by a different method, still make errors in the first trial, whereas group A complete the task with no errors.

Therefore the objectives of this experiment can now be answered.

(1) To highlight specific features of desktop VR/VEs which may be beneficial to training

Desktop VR and VEs can be used to represent a training situation – visually and interactively. However this is not to say that the VE has to be a perfect replica of the real environment. There are obvious differences between the real environment and the virtual environment and these differences should be used to enhance training and not just perceived as limitations. For example, this particular task uses very simple parts which are easily modelled on a computer however this does not mean to say that more complicated parts would not be possible. In fact in some cases it would be better for trainees to experience a less complex environment initially so that they can become comfortable with the procedures of the task without being daunted by the detail, as in flight simulator training. Another advantage of VEs is that it can provide a variety of methods for the trainee to learn about the task and not only the ones that are possible in the real world. The VET in this experiment allows the trainee to see any stage of the process repeatedly as well as allowing 'virtual hands-on' experience.

(2) To train non-experts with the VET application to follow a procedure to assemble a toy car

The VET application was able to successfully reduce the time it would have taken the trainees to reach a certain level of 'competency' as the participants in the control group initially took much more time to complete the first trials. It also appears that more of the information was retained for the second session as the control still took longer times initially to perform the task. With regards to errors, as group B were initially trained using VET with little understanding of the technology or the task, some of the participants found it difficult and made a number of errors (although less than the control group). However group A having initially been trained using the video demo and then the VET, they were able to carry out the task successfully in the second session with no errors. It appears that VET may not be as beneficial to non-experts until they have some understanding of the task to be learned. Therefore the use of VET may be better suited to allow 'practice' of a task rather than learning.

(3) To directly compare the value of VET over another training method.

In the first session, group A who were trained by the video demo. appeared to have better initial times than the participants which were trained by the VET. However these times were not significantly different but it may mean that the video demo. provided more comprehensive training. This view is supported by some of the comments that were made by the participants. They felt that as the VET was an unfamiliar training tool they were required to learn twice as much information i.e. how to use the system as well as learning the task. The VET offered a number of options for the trainee to choose from and as reported in the questionnaires ideally the trainees preferred (or would like to think that they prefer) this method as it gives them freedom to pace and control their learning. However, as Goldstein (1993) found some people are unable to manage their own time and would like to be passive learners rather than active from the start. This needs to be considered when using VET.

By the second session group A who were then trained by the VET appear to be successful and have little problem with using the technology and performing the task. These observations would suggest that a combination of conventional training and VET would greatly benefit a trainee and may be a powerful learning tool. Conventional methods are useful because the trainee already knows how to use the medium to obtain the necessary information. However with a video as the results have shown, it can be uninteresting and not useful when more information is required of a difficult area. While it is possible to 'rewind' for information, it is not possible to see the information from a different viewpoint or to have 'hands-on' practice, all be it *virtual*. VET would add an interactive element to a training tool that could provide the trainee with the flexibility to reinforce what they have learnt in the way they prefer.

In summary, leading on from the first experiment (chapter seven) this second experiment has concentrated on the development of psycho-motor skills with the emphasis on the psychological skills. VET provided a way of allowing the participants to use a variety of methods to view and perform the problem-solving task. The participants were given complete control over their training and all successfully completed the task in times and errors comparable to that of the video demo group. Given the results of these experiments the research questions formulated at the beginning of the study can now be addressed.

9. Discussion

9.1 Introduction

This chapter concludes this work by addressing the research questions formulated at the end of chapter five in the light of the study findings. The questions were:

(1) Is current desktop VR/VEs ready for application to training?

(2) Can VET effectively train basic psycho-motor skills?

(3) How does VET compare to other methods of training?

The section 9.2 discusses the main contributions to the formulation of these research questions outlined in chapters two to five, then section 9.3 discusses how these questions were addressed in chapters six to eight. The key findings are outlined in section 9.4 and then these are discussed in section 9.5 and then recommendations are presented for the way forward for research in this area (section 9.6).

9.2 Formulation of research questions

Virtual environments (VEs) created through virtual reality (VR) technology are still a relatively new idea regardless of the fact that it has been around as a concept since the 1950s (Heilig, 1992). It is a technology that has very much been misunderstood since its emergence into the public domain at the start of the 90's. This has much to do with its unfortunate name - virtual *reality* - and the one view that is often portrayed in the popular press - *headset game systems*. It is therefore understandable why initially some industries have dismissed it as an 'inappropriate' technology to consider for their future needs. Furthermore, there has been much disappointment by industry when faced with the true capabilities and limitations of the current technology. A summary of the main capabilities and limitations of the current VR systems was provided in chapter two and is shown again here in Table 9.1.

Table 9.1: Summary of the main capabilities and limitations of current VR systems

	Capabilities	**Limitations**
Hardware	• a wide variety of platforms to choose • Price is ever-decreasing while capabilities (graphics, sound, videos) are ever-increasing • processing speeds are increasing	• choice of platform places a constraint on the types of components chosen for VR system • processing speeds still impose limitations on all other systems in terms of frame-rates, response times and resolution.
Software	• a wide variety of software to choose from catering for different levels of computer skill • price ranges from 'freeware' to costly but sophisticated software	• software does not integrate particularly well with other computer software like CAD • capabilities of each software can place constraints on the design of the VE
Tracking systems	• choice of mainly four systems at varying costs and abilities	• problems with accuracy, range, response time and interference have to be considered for each type
Visual systems	• wide variety of systems to choose from • low cost options are well-developed (e.g. monitors, projection screens) • high-cost options provide stereo vision and promote immersion and presence (e.g. augmented reality displays, headsets)	• problems to consider include quality of resolution, field-of-view, stereo vs. mono. vision • particular problems with headsets include: cost, weight, side-effects

Auditory systems	• technology has already beed developed for years • use of sound to replace another sense • increase realism of objects and situations • increase immersion and presence	• costly in terms of computer power • difficulty in matching sound directly with visual events • difficulty in spatial location of sounds
Haptic and Input systems	• conventional input devices are well-developed	• limited choice of tactile and force feedback devices • still in development with many problems including: fit, accuracy, type of feedback
Other systems	• motion systems	• little work on smell and taste

Generally there is still a substantial gap between the potential of VEs and the availability of 'enabling' technology. There is a large number of configurations of a VR system and each component has an influence on the others. Of the possible configurations the technical components which make up a desktop VR system - i.e. monitor, personal computer, mouse, joystick, spacemouse - have predominantly been in use for some years and are therefore less expensive and more robust than some other VR systems components. For this reason, and also the wide availability of standard personal computers, desktop VR has real potential to be an affordable first step in the development of applications for much of industry. Therefore this study has focused on the development of applications specifically with desktop VR.

In light of the types of applications being considered for VEs, archive data from the MOVE (Manufacturing Operations in Virtual Environments) programme and other similar studies was examined. The programme was part of an initiative by VIRART (the Virtual Reality Applications Research Team) at the University of Nottingham to examine industrial perceptions of VR, potential VE application areas, likelihood of implementation and user requirements. The work involved many methods of data collection including surveys, case-studies and workshops. While the focus was mainly on manufacturing, data was also collected from twenty other industrial sectors which were categorised into the following groups: computers and software; finance, sales and marketing; services; and manufacturing. The information was revised in respect of the other industries and disseminated *in "Industrial Applications of Virtual Reality: opportunities and limitations*" (Wilson et al, 1996), and is summarised in chapter three of this work. The application areas of interest to industry, highlighted by MOVE and confirmed by other studies (e.g. NRC, 1995; HITLab, 1997; Hand, 1997) are mainly: military applications; games and entertainment; medicine and healthcare; design, manufacturing and engineering; sales and marketing; education and training. Of these areas 'training' is seen as likely to be an 'early' application of VEs (NRC, 1995), especially as it is of interest to most industrial sectors.

Furthermore it was also possible to identify the main barriers to implementation of VEs. Apart from cost, the barriers appear to be the current technological limitations, usability issues, the lack of 'working' applications and therefore the lack of any evidence of added value. While research work in each of these areas is on-going, it was felt that these barriers may be addressed and reduced through some form of structured guidance for developing and evaluating VEs (Wilson et al, 1996; D'Cruz et al, 1996). In particular, as there is certainly great interest in evidence of the potential of VEs, the area of evaluation has provided a further focus for this work Therefore the overall aim of this research was to:

Investigate and further the development and evaluation process of virtual environments in the field of industrial training.

9.3 Addressing the research questions

This overall aim (and also the particular research questions developed after background research) were addressed through the following objectives stated at the beginning of this study (section 1.3), namely to:

1. *Critically review the capabilities and limitations of current VR systems and VEs for their impact on training applications*
2. *Identify the types of training applications currently being pursued by industry using VEs, highlight the findings and major problems that need addressing;*
3. *Investigate existing theories and methodologies of training and evaluation for their recommendations on developing and evaluating effective VET applications*
4. *Suggest a structured framework in which to develop and evaluate effective VET applications;*
5. *Develop a VET application based on the needs of industry and explore the costs and benefits against other forms of training through experimental work*
6. *Provide recommendations for future research.*

The first objective - to provide a critical review of the current VR systems and VE applications - was delivered in chapter two and chapter three as discussed and summarised in Table 9.1. The second objective was delivered in chapter four. Through the literature, conferences and contacts with developers and suppliers of VR systems the better known case-studies in the area of virtual environment training (VET) were identified. In most cases the developers or clients were contacted and some were visited in order to, gather further information, in particular about evaluation of their VET. Generally the types of VET applications that are being developed are as follows:

- *Familiarisation with a product/process/place;*
- *Rehearsal of procedures for repair and maintenance; and*

- *Problem-solving/trouble shooting scenarios.*

They tend to be 'proofs-of-concept' and therefore little evaluation has been performed. Of the evaluation studies that have been carried out, the main objectives have been to measure the effectiveness of VET and transfer-of-training but these evaluations. However most of these have been confounded by the many uncontrolled variables which may have accounted for learning and performance effects. Generally the conclusions have stated difficulty in determining value due to usability problems. These problems have occurred as they have often used headset systems which have the added disadvantage of causing some people undesirable side-effects that effect performance (Cobb et al, 1998). This confirmed that it would be of more use to the immediate needs of industry if this study evaluated desktop VR for VET. Desktop VR has less of the side effects associated with other types of VR systems and is the likely choice of system in the short-term because of cost (see chapter two). Furthermore in spite of its popularity, there is little evaluation work that has been carried out in the area of desktop VR.

The third objective - to investigate existing theories and methodologies on evaluation and training - was delivered in chapter five. The information available in the literature on training and evaluation is vast and in total beyond the scope of this study. An indication and summary of the VR/VE relevant information on the training development process and especially the evaluation of training has been provided. Also lessons that can be learnt from similar technologies to VR, such as computer-based-training (CBT) and simulators, are included. On the whole the area can provide many guidelines to enable the development of effective VE training. In terms of evaluation many of the problems which face the evaluator of a CBT or simulator will also effect the evaluation of VET. Patrick (1992) states that the problems arise mainly due to the differences in design, presentation and use of the various training methods which makes it difficult to perform fair assessments. However, though it is beneficial to learn from similar technologies, even in the field of CBT there has been little evaluation work carried out (Patrick, 1992; Goldstein, 1993). The largest study is still the one carried out by the U.S. military over 20 years ago (Orlansky and String, 1979). Technology has obviously moved on since then as well as, how people like to learn. Therefore it is important to consider existing theories in line with current trends.

In light of this, it was felt necessary to develop a new framework with grounding from existing models on developing training programmes and further concepts particular to VE development (begun in the framework suggested in chapter three). This new approach was presented in chapter six and fulfils the fourth objective - to suggest a framework in which to develop and evaluate VET applications. Each stage of the development process needs addressing and this will be discussed further in section 9.5. Organisations need to have some kind of 'VE selection criteria' for the proposal stage in order to understand the potential of VETs and examine it against their requirements. Other selection proposals are discussed (Dorrington and Elliot-Square, 1995; Adams and RTI, 1997) and then the author's own suggestion based on the attributes of VEs, experience within VIRART and previous models. With regards to specifying the VET, guidelines are offered based on work by Bee and Bee (1994) but made specific to VET by the author. Guidelines for VE building present difficulties, because of the variety of software available for building VEs which work in different ways. However there are a number of recognised stages – information gathering, development of design, modelling of the VE, user's interaction and review/modification sessions (Eastgate et al, 1996; 1997). Finally evaluation also has a number of stages. Formative evaluation (Tessmer, 1993) examines 'process' information i.e. information to identify revisions needed in the design of the application; and involves methods like self-evaluation; review sessions with proposers; expert and non-expert reviews; and small group reviews. Summative evaluation examines 'outcome' information i.e. information to assess the success of the application for its intended purpose. It involves information at different levels: reactions;

learning/immediate outcomes; job behaviour/intermediate outcomes; results/ultimate outcomes/organisational outcomes; and ultimate value (Hamblin, 1974; Kirkpatrick, 1967;1996; Warr et al, 1978). Further to this, the author has developed a VR/VE assessment tool which examines each component of the VR system and VE to quantify the user's opinion. Details of this can be found in Appendix (II).

Within this framework, the fifth objective - to develop a VET application based on the needs of industry and explore the costs and benefits against other forms of training - was achieved and delivered in chapters seven and eight. VET applications were proposed, specified, built and evaluated against conditions of no training and training by alternative methods. The first experiment (chapter seven) involved training to replace a network card in a computer. The second experiment (chapter eight) involved an assembly task using Lego. The results of these experiments are discussed in the next section 9.4.

The deliverables of the last objective – future recommendations - are considered in section 9.6.

9.4 Key Findings

Before discussing the research questions outlined at the beginning of this chapter (section 9.1), a summary of the results and key findings of the two experiments is as follows.

The first experiment (chapter seven) involved training to replace a network card in a computer. It consisted of one session with three conditions: (1) control that received no training; (2) a group trained by a video; and (3) a group trained by a VET.

The results are summarised in Table 9.2.

Table 9.2: Summary of results from experiment one

Level	CONTROL	Video	VET
Reactions	• found the task difficult • made expected errors	• clear • straightforward • uninteresting • boring	• interactive • self-pacing • enjoyable • interesting • usability difficulties with spacemouse
Learning	Mean = 345s SD = 168.61s	Mean = 230s SD = 66.05s 34.3% time savings	Mean = 230s SD = 107.96s 35.3% time savings
Cost benefits/ cost effectiveness		~ 2 days £550/ per day +	~ 2 months £350 - £800 per day + Novel approach

Reactions of the participants were different depending on which group they belonged to. The control group felt the task was difficult and would have preferred some training. They also made the expected errors identified through the pilot studies. The video group found the training method clear and straightforward but unfortunately uninteresting and boring. In comparison the VET group thought it was interactive, self-paced, enjoyable and interesting, but many of them had usability problems with the spacemouse. As for learning outcomes, the times of the subsequent performance task were not statistically significant between the groups although time savings of 34.3% (video) and 35.3% (VET) were achieved. In terms of cost benefits/cost effectiveness, the development cost of a VET is very much greater than that of a professional video. The main benefit of the VET over the video is that it is a novel approach which sustains interest. Justification for the cost can only be if the task is highly important and the enjoyment of the trainee is critical. This task was chosen as it was a basic psycho-motor task similar to the tasks already being explored by industry (chapter four). However through more detailed observation of the performance task, it was revealed that the most difficult steps were related to motor skills performance, i.e. use of the screwdriver and replacing the outer casing on the computer. The VET could not adequately represent the motor skills experience required given the configuration of the desktop system.

To reduce the effects of differences in motor ability, the next experiment, experiment two (chapter eight) used Lego models. The models use simple building blocks that easily attach and detach without the need of specialist tools. Also in order to allow the participants an opportunity to directly compare the two methods of training, they participated in two sessions. During each session the participants were required to assemble a toy car from memory and therefore the emphasis was on developing the 'cognitive' part of this psycho-motor task.

The results of experiment two are summarised in Table 9.3 (overleaf). Similar to the reaction results of experiment one, the control group made many errors and would have liked to have had some training. The participants thought that the video demo was effective but uninteresting and limiting as the environment was only viewed from one angle. In contrast they thought that the VET was interesting, highly interactive and, when given the choice, the majority (23/24) preferred the VET.

Table 9.3: Summary of the results from experiment two

Levels	**CONTROL**	**Video demo.**	**VET**
Reactions	• made many errors • would have liked some training	• effective • uninteresting • limiting	• interesting • highly interactive • preferred choice
Learning outcomes	**SESSION ONE**		
	Significant difference in times between: Control and Video demo in trial 1, 4 & 5 Control and VET in trial 1 & 4		

	NO DIFFERENCE IN TIME BETWEEN Video demo and VET in any trial Significant difference in errors between: Control and video demo in trial 1 & 2 Control and VET in trial 1 NO DIFFERENCE IN ERRORS BETWEEN Video demo and VET in any trial
	SESSION TWO
	Significant difference in times between: Control and VET in trial 1 & 2 Control and Video demo in trial 1 NO DIFFERENCE IN TIME BETWEEN Video demo and VET in any trial Significant difference in errors between: Control and VET in trial 1 NO DIFFERENCE IN ERRORS BETWEEN Video demo and VET in any trial

In terms of learning outcomes, the significant differences are summarised in terms of groups and trials. The results showed that session one had a considerable 'learning' effect on session two, therefore the performance times to complete the tasks for each session could not be statistically compared however they did highlight some interesting information. Generally the group that was trained by the VET - group B - in the first session, had difficulties with the technology. They had to use as much of the training time in understanding how to effectively interact with the VET as learning how to perform the task. As a result, although not significant, they still took slightly longer and made more errors than the video demo group - group A.

The VET group though, was significantly different to the control group showing that certainly some value was gained from the VET. The conclusion was that the both training methods were better than no training and the VET was just as effective as the video demo.

In the second session, group B having previously been trained by the VET were then trained by the video demo. Even though they had awareness of what was likely to be involved in the new task, they still had difficulties in performance. This could either be accounted for by individual differences or may suggest that the video demo training was inadequate. In the first session the VET had allowed them to interactively learn the task. While this had proved difficult because it was unfamiliar, when presented with the video demo, they appeared to have a greater disadvantage. The video demo may not have conveyed as much information as was required, and the participants were unable to specifically focus on difficult steps. In comparison, group A who were trained initially by the video demo and then by the VET, appeared to have difficulties in the first session and then successfully completed the task with no errors, in the second session. This again could be due to individual differences or possibly after the first session, they were aware of what was involved in the new task and so used the VET to specifically concentrate on difficult steps. This implies that maybe the combination of VET and conventional forms of training is a more powerful training tool, rather than VET in isolation.

9.5 Conclusions

Given the key findings discussed above in the previous section (9.4) the research questions set at the beginning of the chapter (section 9.1) can now be answered.

(1) Is current desktop VR/VEs ready for application to training?

The desktop VR systems used in experiment one and two were certainly capable of training the participants however they also highlighted potential areas of difficulty. In particular the limitations of current input devices. Input devices are the devices that allow the user to interact with a VE. They must be simple to use and effective for performing the given task. In experiment one the input device chosen was a spacemouse. This was because it gave the user six-degrees-of-freedom to navigate freely around the VE. The alternative would be the combination of keyboard, mouse or joystick which may prove too complicated. The spacemouse is also a device which is commonly used with a desktop VR system. However it presented most of the participants with difficulty which meant that much time was spent in learning how to use the spacemouse rather than learning how to perform the task. This was eliminated to some extent with in experiment two by only providing the mouse as an input device. However the participants still had usability problems, mainly in understanding how to effectively use the system in the given time limit. Therefore it seems that while desktop VR systems appear ready, the potential users are not. It is essential that a trainee spends more of their time in learning how to perform the task than how to perform the training. In the current economic climate, there is an emphasis on decreasing the time it takes to learn while increasing the number of people who receive training (Training Agency, 1989; Bee and Bee, 1994). Unless there is some form of in-built training or 'help' function found in many other systems to train users how to effectively use their VR system, VR technology may get left behind.

Therefore compared to the currently available VR technology, desktop VR appears to be the most ready for application to training. This is mainly because, it requires minimal investment cost comparably and the current configuration has less side effects associated with other technologies. Whether desktop VR *should* be used for training is a different question that has been partly considered in chapter six. VE selection criteria were designed by the author to suggest when it may be appropriate to consider the use of VEs for training. It has already been highlighted that the technology has limitations which need to be addressed however it also has capabilities which have the potential to add value to conventional methods. What must always be remembered is that it is the nature of the training application which will determine whether VET should or should not be used.

(2) Can VET effectively train basic psycho-motor skills?

A number of previous studies and experiment one in this study (chapter seven) have shown the potential difficulties with tasks where motor skills are important for performing the task. As discussed above, current input devices available for desktop VR can provide certain types of movement through the environment but interaction is still very much limited. There is some glove technology but little that is available that provides the user with the physical experience of using a tool like a screwdriver. Therefore, VR, and in particular desktop, is restricted in providing motor skills experience. In terms of psychological/cognitive skills, some research has already proved benefits of VEs for training navigational skills (Regian et al, 1992; Witmer, 1996). Certainly VEs provide good visual and aural cues for problem solving and familiarisation tasks but any further senses than these, is difficult. So VEs can train psycho-motor skills but mainly where the emphasis is on cognitive skills rather than motor skills. VEs can certainly provide a visual representation for motor skills but is restricted when providing the physical experience. If this is required then possibly the combination of VE and a more 'hands-on' approach may be necessary. VEs could be used to practice the mental processes of the task before actual hands-on experience. This may reduce the amount of time required on the real equipment and possibly the amount of errors. Thus proving to be highly beneficial and economical for training psycho-motor skills.

(3) How does VET compare to other methods of training?

Experiments one and two both compared VET with a more conventional approach to training such tasks – passive demonstration. Generally the participants had a more positive reaction towards the VET probably due to its novelty and the interactivity offered by the application. The participants reported that they enjoyed being involved in their own training and preferred the VET to the video. However some participants, particularly in experiment one, had usability *difficulties with the system. Also some felt that because this was an unfamiliar technology they found themselves concentrating on learning to use the system rather than learning the task. As discussed before this area needs to be addressed as the situation may change once the participant becomes an experienced user.*

In contrast the participants found the passive demonstration uninteresting but they were all familiar and comfortable with using the method and found it generally effective. Therefore in both experiments the VET proved to be just as good as conventional methods. Although experiment two seemed to show that the combination of conventional training and VET has the potential to be a powerful and useful tool for the future. Especially if there is difficulty in providing real hands-on training. The potential of current VET is in at least providing the mental processes of the interaction of the task. As technology develops and our experience of the technology improves, it is likely that eventually VET will also be able to provide experience of physical processes too.

In terms of cost, unfortunately the development and running costs of VET are still relatively high compared to conventional methods. The main cost appears to be in development time. The lack of understanding of the capabilities and limitations of the technology and lack of any real guidance, implies that much of the time is involved in learning how to build an effective VE. Furthermore it is difficult for a developer to predict exactly what is or is not possible and how long it is likely to take unless the proposer of the application has fully specified exactly what is required and expected. Generally developers have to be aware of 'scope creep' (I/S Analyzer, 1997), where the proposer continues to make the project larger then originally planned throughout the development process. it is difficult to know exactly when to stop developing the application as a VE provides 'infinite' possibilities. However the longer the development process takes the more costly the project becomes.

Furthermore there is still little known about the actual effects of VET. For example even though it has the potential to present the training environment in many different ways there is little evidence that this actually enhances the training for the trainee. In some cases it may prove to confuse them by offering too much choice. As was shown in the research by Hartley (1985) some trainees cannot set their own pace of learning and initially need some instruction. This further confirms the use of VET in combination with more conventional methods of training.

9.6 Recommendations for Future Research

As this study has suggested each stage of the development process outlined in chapter six and illustrated again here in Figure 9.1 (overleaf) need to be addressed. Each stage in turn is discussed as follows:

9.6.1 Proposal of VET applications

The question of *should* VEs be used for the application is an important issue. There is still little that is understood about the capabilities of systems but the nature of VEs is better known and not likely to change. Therefore it is possible to develop guidelines for proposing VET applications, as what is required is detailed examination of the attributes of VEs and how these effect the attributes of a training application. This was begun in chapter six.

The author suggested some form of selection tool but this requires further development and testing. Such guidelines would provide a highly useful initial step for industry in their consideration of VEs. There is little point in industry implementing a technology that does not suit their current needs as it leads to costly disappointment. Alternatively if VEs can provide a unique solution to a particular need then it would be advantageous to have guidelines to provide evidence of this. Therefore this is a highly important area to address.

9.6.2 Specification of VET applications

There is currently no standard ways of specifying VET applications. This generally involves many meetings between the proposers and the VE developers where eventually the application evolves. This process can be made more constructive by providing some generic guidelines. The training field already has a number of suggested guidelines (Bee and Bee, 1995) but these have to be reviewed in light of specific requirements of the VE developer. The author again has suggested an initial set of guidelines but these need to be further developed and tested. Such guidelines could greatly reduce development time and costly errors made through misunderstandings.

9.6.3 Building of VET applications

There are currently no standard guidelines for building VET applications. Such guidelines are difficult because of the differences in the developer's kits. However there are common issues in all VEs as suggested in chapter six and including: the best method of representing the user; which viewpoints to include; degree of interaction; and cues for interaction (Eastgate et al, 1997). Further work is required to minimise development time but also to provide methods for effective VEs in terms of utility and usability.

9.6.4 Evaluation of VET applications

Finally with regards to evaluation, this is the area that has the most interest but little actual work. One of the reasons has been the immaturity of the technology which has meant that there have been few case-studies to actually perform evaluation studies on. However with evaluation studies more can be understood about VR and VEs which is likely to encourage further development. What is required in this area is as follows:

- Development of standard VR and VE evaluation techniques and methodologies (as suggested by the author in Appendix II) in order that results can be compared;
- Further evaluation studies on systems that are the most likely to be used by industry, for example desktop VR, in order to encourage implementation;
- Closer examination of how the different components of a VR system interact. For example, for the system to have a powerful headset, it requires a powerful computer. Therefore some form of classification of the different components would be useful for industry choosing the appropriate system;
- Closer examination of how the different components of a VE affect the user. For example, are the colours too bright, are the objects too simple/complicated etc.
- Closer examination of how the different components of the VR system affect the user. For example, are there usability difficulties with the input devices, etc. This should lead on to the development of training programmes to aid the understanding of how to use VR systems and VEs effectively.
- Comparison of VET with lots of other forms of training. For example, lecture-based training, interactive CD-ROMs, simulation, etc.
- Comparison of the combination of VET and conventional training against just the use of conventional training.
- Wider dissemination of information on evaluation studies so that the research community and industry can continue to make progress.

In summary, this study has gone part of the way toward investigating the potential of VEs for training applications. As the recommendations for future research suggest, such application is still in its infancy. This is because the technology is still not yet fully matured, applications are still uncertain and untested and methods for standardising the development and evaluation process are only now being considered. Such methods would greatly reduce the development time and give confidence in assessment, and therefore may make implementation of VET more feasible. The future of VET may be dependent on such developments.