CASS SCHOOL OF EDUCATION

COMPETENCY BASED ASSESSMENT USING VIRTUAL REALITY (VERT): IS IT A REALISTIC POSIBILITY?

DAVID M. FLINTON

A Thesis submitted in partial fulfillment of the requirements for the degree of a professional Doctorate in Education

October 2015

Abstract

The education of the radiography profession is based within higher education establishments, yet a critical part of all radiography programmes is the clinical component where students learn the practical skills of the profession. Assessments therefore not only have to assess a student's knowledge, but also their clinical competence and core skills in line with both Health and Care Professions Council and the Society and College of Radiographers requirements. This timely thesis examines the possibility of using the Virtual Environment for RadioTherapy (VERT) as an assessment tool to evaluate a student's competence so giving the advantage of a standard assessment and relieving time pressures in the clinical department.

A mixed methods approach was taken which can be described as a Quantitative Qualitative design with the emphasis being on the Quantitative element; a so called QUAN → qual design. The quantitative evaluation compared two simulations, one in the virtual reality environment and another in the department using a real treatment machine. Students were asked to perform two electron setups in each simulation; the order being randomly decided and so the study would be described as a randomised cross-over design. Following this, qualitative data was collected in student focus groups to explore student perspectives in more depth.

Findings indicated that the performance between the two simulators was significantly different, $p < 0{\cdot}001$; the virtual simulation scoring significantly lower than the hospital based simulation overall and in virtually all parameters being assessed. Thematic analysis of the qualitative data supported this finding and identified 4 main themes; equipment use, a lack of reality, learning opportunities and assessment of competence. One other sub-theme identified for reality was that of the environment and senses.

Declaration

I hereby declare that, except where explicit attribution is made, the work presented in this thesis is entirely my own.

Word count (exclusive of appendices, the list of references and bibliographies, but including footnotes, endnotes, glossary, maps, diagrams and tables): 54,902 words.

D Flinton

David Flinton

Table of Contents

List of Figures

List of Tables

List of Abbreviations/Glossary

app	This abbreviation is used in two forms. The first use is an abbreviation for a software application used on an IPad. The second use of app is found in the student quotes where app is a shortened from of apposition and is used in radiotherapy to describe bringing the applicator close to and parallel to the skin, "skin app".
AR	Augmented Reality. *A real-world environment that is supplemented or overlaid with computer-generated sensory input(s).*
ARRT	American Registry of Radiological Technologists. *Credential organisation for radiographers within the USA whose function is similar to that of the COR, promoting high quality patient care in the field of radiography.*
AV	Augmented Virtuality. *A virtual world with a degree of reality in it.*
COR	College of Radiographers *A charitable subsidiary of the Society of Radiographers. Its main objectives are directed towards education and research in the support of the practice of radiography.*
CPD	Continuing professional development. *Education of health care professionals following completion of formal training. It is an expectation of all radiographers and an HCPC requirement.*
CPF	Career Progression Framework *Framework initially launched to progress radiography careers that introduced four levels of clinical practice ranging from assistant practitioner to consultant practitioner.*
DCR	Diploma of the College of Radiographers *Registerable qualification for radiographers that was replaced by an undergraduate degree around 1992/3.*
DICOM	Digital Imaging and Communications in Medicine *A standard file format developed by The American College of Radiology together with the National Electrical Manufacturers to support the communication of digital image information, regardless of device manufacturer.*
DoH	Department of Health *A ministerial department concerned with health and care.*

$_{d}$max — Dose Maximum
The maximum dose point of a beam or treatment. It can be talked about in relative terms when it will be quoted in Gray, but is more often normalized to 100%.

EBRT — External Beam Radiotherapy
T*he most common form of radiotherapy consisting of a beam of radiation directed towards the patient from a distance.*

ERCP — Endoscopic Retrograde Cholangiopancreatography
A surgical technique to diagnose and treat conditions of the bile ducts and main pancreatic duct.

FSD — Focus to Skin Distance
Distance between the focal spot (where x-rays are produced) of the x-ray unit to the skin surface of the patient.

GLM — Generalised Linear Model
A flexible generalization of ordinary linear regression that allows for the transformation of the expected response or as a nonlinear regression model for the response.

HCPC — Health and Care Professions Council
Regulatory body for health professions including radiographers. Anyone using the title radiographer must be registered with the council in order to practise in the UK.

HEI(s) — Higher Education Institute(s)
Usually Universities that can either award a bachelor's degree or, provides not less than 1 year of training towards gainful employment or, offers a vocational program that provides training for gainful employment and has been in existence for at least two years.

H0 — Null hypothesis
A general statement or default position that there is no statistical difference/relationship between two measured phenomena.

ICRU — International Commission on Radiation Units & Measurements
A committee that exists to develop and promote internationally accepted recommendations in the field of radiography, including radiation related quantities, units, terminology and procedures.

ITQ — Immersive Tendencies Questionnaire
Questionnaire developed by Witmer and Singer to measure presence in virtual environments.

KSF — Knowledge Skills Framework
A framework tool to identify the knowledge, skills and learning and development that NHS staff need to do their job and undertake personal development planning.

LCS	Liquid Crystal Shutter *Liquid Crystal Shutter glasses consist of a liquid crystal layer that can turn opaque so as to alternately block one eye. This occurs in synchronisation with the refresh rate of the screen can give the impression to the brain that the person is looking at a 3D image.*
LINAC	Linear accelerator *Treatment machine that can produce high energy x-rays and electrons by using radiofrequency waves to accelerate electrons down a linear evacuated tube. Used almost exclusively in the treatment of cancer.*
LMPA	Low Melting Point Alloy *A range of alloys that can be used to create customisable blocks and cut-outs which are used to shield parts of the radiation beam to protect parts of the body from radiation.*
MAR	Missing at Random *Data values that are missing that are related to a particular variable, but are not related to the value of the variable that has missing data.*
MCAR	Missing Completely at Random *Data values that are missing that are independent both of observable variables and of unobservable parameters of interest, and occur entirely at random.*
MNAR	Missing Not at Random *A values in a data set that is missing for a specific reason (non-random).*
NHS	National Health Service *A publicly funded health care system that was originally launched in 1948. It is often used to encompass the four publicly funded health care systems in each country of the United Kingdom, but officially it refers just to England.*
NRAG	National Radiotherapy Advisory Group *A group set up in 2004 to advise the Government on the current position of radiotherapy services in England in order to ensure resources are deployed to best effect and to advise on future directions.*
PTSD	Post-traumatic Stress Disorder *This is a mental health condition that is usually triggered by either experiencing or witnessing a terrifying event that usually involves physical harm or the threat of physical harm. Symptoms are typically classed into three types, re-experience, avoidance and hyper-arousal.*
PIXY®	A type of anthropomorphic phantom. *PIXY is 156cm tall female weighing 48kg. The phantom is made of tissue-equivalent materials and has life-like articulations.*

PQ	Presence Questionnaire *A questionnaire designed to measure presence in virtual environments.*
RVC	Reality-Virtuality Continuum *Framework suggested by Milgram to look at the concepts of reality and virtual.*
SCOR	Society and College of Radiographers *Trade union and professional body representing radiographers in the UK.*
SOR	Society of Radiographers *The professional body and trade union that represents radiographers in the UK.*
TLRP	The Teaching and Learning Research Programme *The UK's largest educational research programme that coordinated research and investment in research between 2000 and 2012.*
t(d)	Welch's t-test *An adaptation of the standard t-test, which is more reliable when the two samples have unequal variance.*
TURP	Transurethral Resection of the Prostate *A surgical procedure under anaesthetic that involves inserting a resectoscope into and up the urethra to cut away section(s) of the prostate gland.*
VE	Virtual Environment *A computer-generated, three-dimensional representation of a real life setting which can be explored and interacted with.*
VERT	Virtual Environment for Radiotherapy *A virtual environment consisting of a radiotherapy treatment room, treatment machine and patient.*
VR	Virtual Reality *Used synonymously with VE. A computer-generated environment that simulates physical presence in real or imaginary worlds.*

Acknowledgements

Many people have contributed to my learning experience during the process of my studies at UEL. I am thankful to my two Directors of Studies, Dr. J. Truschel and Professor J. Preston, for their insight, thought provoking questions and guidance for my thesis.

In addition, I would like to acknowledge my PhD thesis supervisor Dr Chadderton for her help and reassurance at various stages of the work.

I would also like to thank my work colleagues in the Department of Radiography, Mrs P. Cherry, Miss C. Raymond-Barker, Mr R. Thorne and Mr R. Khine. They have kindly taken on extra work themselves in order to allow me time to undertake the necessary data collection and subsequent write up. Perhaps most importantly they were always there to bounce ideas off and offer encouragement and support, especially when times were bad.

Dr James Ward, Managing Director, Head of Research and Development and Founder of VERT who found the time to help with the electron field graphic overlays for the VERT phantom.

Finally and perhaps most importantly the participants as without them the work could not have been done. A big thank you to the students of the 2010, 2011 and 2012 radiotherapy cohorts at City University London.

Dedication

I am especially indebted to my wife, Helen and my two sons Alan and Peter for allowing me to put part of my life on hold during the last few busy years whilst I struggled with the development of this thesis, which was largely for my own personal development.

Finally in memory of Dave Wood, Therapeutic Radiographer. A student in the 2012 cohort who was tragically killed in a road traffic accident in 2015 shortly after qualifying.

"Well some say life will beat you down.
Break your heart, steal your crown.
So I've started out for God knows where.
I guess I'll know when I get there.
I'm learning to fly."

Tom Petty (1991)

Chapter 1 - Introduction

1.1 Background

Recently there has been a paradigm shift in higher education with students being increasingly placed at the centre of the learning process. This view of learning is diametrically opposed to the traditional approach that tended to consider students as passive receptors of information. Students would typically sit in lecture theatres and not be an active participant in the learning process, (TLRP 2010). The didactic lectures would be the students' main source of knowledge, there being little additional material to support learning, and students were rarely expected to contribute, ask questions or challenge the expertise of the academic, (McCarthy & Anderson, 2000).

With the move towards a more student centred approach to learning educators are facing the challenge of finding ways of putting the student at the centre of the learning experience. Students have to be engaged in order that the student should learn by doing, and it is in this respect that technology may help. One recent technology that has been identified as potential enhancement pedagogy is that of Virtual Reality (VR), a computer technology that was developed in the early sixties and allows the creation of a Virtual Environment (VE) which users can interact with in real time in an autonomous way.

Simulation has been a widely method of training in both the medical and para-medical fields for a number of years and is an important feature of many programmes particularly for skills training, although they are also used to measure competency and look at interpersonal skills, (Kalaniti & Campbell 2015). VR represents a substantial evolution in simulation and potentially offers medical and para-medical educators a number of advantages as a method of simulation over the more traditional methods. According to Moore (1995), the significant key benefit offered by VR is its ability to provide immersion of the user allowing the user to experience, explore and interact with digital knowledge first hand to help develop understanding of the real world or concepts related to it. Finally, it may bring information to the user which is not normally available in traditional educational settings, enabling users to experience life and objects as never before, as users may step into and interact with minutia such as manipulating atoms and molecules or compress large objects with users being able to fly through the solar system and explore planets.

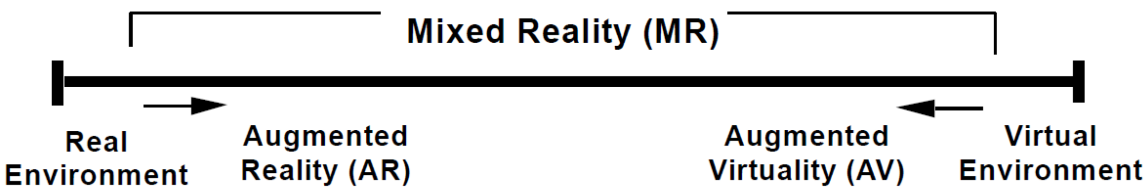
Supporting this stance, Bell & Fogler (1995) stated that VR addressed each of the dimensions of those learning styles proposed by Felder & Silverman (1988), i.e. sensory/intuitive, visual/auditory, inductive/deductive, active/reflexive and sequentially/ globally (*ibid*.: 675). On the sensory/intuitive scale, VR can provide a tangible representation of abstract concepts. VR is highly visual and therefore meets the needs on the visual/auditory scale, and can provide non-verbal auditory stimulus which is important to the realism of the overall experience and can also be used to provide educational sound cues, such as the sound of bonds breaking as atoms interact. On the inductive/deductive scale, VR can provide a medium for exploration and learning by observational experience. Virtual reality is highly active and immersive, the main value of VR is that the subjects are inside the simulation and so are active participants. The final component of Felder and Silverman's learning styles is the sequential/global scale. VR can address the needs of the global learners by showing the inter-relationships of the real and abstract concepts, as VR can suspend or work within the physical reality so allowing the learner to work within the big picture.

However, as with all technologies, there are potential disadvantages. Whilst VR systems are getting cheaper, they are by no means cheap and some of the devices can be clumsy for the user which can detract from the immersive experience. Other issues that can affect the use of VR and the impression of reality are immersion, presence, cybersickness, lag, and restriction of view. Dalgarno & Lee (2012) broadly support the concept that benefits are offered by VR, but also counselled caution as, although their study provided a strong support for the idea that benefits of VR occurred, the results suggested that the links between the learning tasks and learning reported benefits from the perspective of the user were minimal and not statistically significant. Although presence, the sense of being in the virtual world is considered a cornerstone of virtual reality, Hodges *et al*., (1994), and Whitelock *et al*. (2000) express reservations for fully immersive environments as their findings suggest that a subject's performance can be negatively affected by cognitive overload.

It must be remembered that VR is still considered a relatively new technology having only being in mainstream use for approximately 25 years. Moreover, although the performance and abilities of VR systems have increased immensely in this time, the potential for VR systems to seamlessly integrate imagination and interaction with computer generated reality shown in science fiction television programmes such as Star Trek: The Next Generation; Red Dwarf and films like The Matrix has not yet been achieved. VR can still therefore be considered to be in its infancy and is still principally seen as a training tool, it being primarily associated with skills development, rather than gaining knowledge (Walsh *et al*., 2010). Proponents might consider this view of virtual reality use as being rather limited and other uses are currently being investigated by various researchers. At present the use of VR in general, and especially these newer areas of study, must be considered as being at an early stage of development rather than an established educational tool.

VR is one extreme on the reality-virtuality continuum (RVC) all of which may have uses in education. The RVC was first proposed by Milgram *et al*. (1994) and is shown below in Figure 1.1. It has on the left one extreme of an environment consisting only of real objects, and on the right the opposing extreme of virtual environments consisting exclusively of virtual objects, such as those described above. In between these two extremes we have mixed reality, sometimes referred to hybrid reality, realities that consist of a mixture of real and virtual objects. Mixed reality itself can be further divided into Augmented Reality (AR) and Augmented Virtuality (AV). Both AR and AV allow real and virtual objects to coexist in the same space and be interacted with in real time, the difference being that AR is built around a real environment, adding virtual media to the environment, (Ternier *et al*. 2012) whereas AV is a virtual world with a degree of reality present in it.

Figure 1.1 Simplified representation of a RVC.



Milgram *et al*. (1994: 283)

The use of virtual, whether it is fully virtual reality or mixed reality offers diverse opportunities in education at all levels. As with VR, mixed reality is being put to a variety of uses including education, Bower *et al*. (2014: 1) stating that "Augmented Reality is poised to profoundly transform Education as we know it." The main advantage offered by all these methods is that they are able to root the learning in experience. Because of this virtual and mixed reality environments have specific opportunities to be beneficial for children with both mental and physical disabilities, VR minimising the effects of a disability, allowing learning through social participation also potentially improving the child's quality of life, (McComas, Pivik, & Laflamme, 1998). This method therefore suits itself to both primary education and medical and paramedical fields where experiential learning is important, the reported main advantages of such systems being learning and motivational gains, (Bacca *et al*. 2014). As with VR it is the immersive nature of the mixed reality systems that is cited as a major advantage, enhancing the student's educational experience, (Lindgren & Johnson-Glenberg, 2013; Kamphuis *et al*. 2014).

Outside of the education forum virtual reality is also still being used in the military as a training aid. Its use however, has also expanded into therapies, particularly psychological conditions such as anxiety disorders for example arachnophobia, pteromerhanophobia and post-traumatic stress disorder (PTSD), where patients are exposed to the triggers for their condition, which they gradually become acclimatized to. This has the effect of decreasing their symptoms and enabling them to cope to better when challenged by their anxiety trigger, and in the case of PTSD a change in their fear response, (Opriş *et al*., 2012).

1.2 VERT

The VR technology that this study considers is called VERT, (Virtual Environment for Radiotherapy) and is an educational based VR system that has been used for radiotherapy training since 2007. The VERT system is available in a number of variants designated by the manufacturers as desktop, seminar and immersive. Desktop, as the name implies, is designed to operate directly on a computer, whereas the seminar system utilises front projection to project an interactive model onto a screen. The immersive model uses a back projection system that allows the user to walk around a virtual radiotherapy treatment room which can be visualised in either 2 or 3 dimensions. It also has the ability to be used in "tracking mode" which alters the viewer's position within the virtual world automatically as they change their position within the "real" room. These variants offer increasing immersion within the virtual world although at the expense of increasing cost.

Initially only two institutions, University Aarhus Hospital (Denmark) and Birmingham City University had invested in the VERT system when the UK government launched the Cancer Reform Strategy (2007). Central to the strategy was the consideration of the predicted demands on the existing service. Currently it is estimated that approximately 1 in 3 people within the UK will develop a cancer sometime during their life and, although modellers believe there will be no change in the age standardised incidence rate of cancers, there will be a change in the actual number of cancers being reported. This increase is due to the aging population and it is predicted that, over a twenty year period covering 2010 to 2030, there will be a 55% and 35% increase in male and female cancers respectively (Mistry *et al*., 2011). Subsequent to this prediction, the Government allocated £500 million for new and replacement equipment including 167 new linear accelerators (LINACs). In order to prepare the necessary radiographers to staff the new units - and to address the shortfall in radiographers at the time a further five million pounds of capital funding was allocated to provide both Higher Education Institutes (HEIs) and hospitals with radiotherapy departments with the VERT system. The purchase of the VERT systems detailed in the Government report was based on the recommendations of the earlier UK National Radiotherapy Advisory Group (NRAG) report (2007) that highlighted the need to introduce VERT as a means to prepare more radiographers. VERT's role would be dual fold: firstly, it was proposed that it would combat the high student attrition on radiotherapy programmes by improving the learning experience; and, secondly, it would increase training capacity.

The idea that direct investment in technology will give social, educational and economic improvements such as those proposed by the NRAG report has been criticised as reductionist in nature, (Chandler, 1995). Problems such as those raised by the NRAG report of attrition and quality are being condensed to just a technological issue which can be corrected with direct investment in technology, a line of reasoning that is founded in technological determinism (Gunkel, 2003). Noeth & Volkov (2004), however, stated that the major criterion for technological implementation is not the technology itself, but whether the technology and its application improve teaching and learning. Although the use of VR in education has potential benefits, as identified earlier, the VERT system in which the Government invested was untested at the time for the goals for which it was purchased.

1.3 Simulation & Assessment

Simulation and simulators such as VERT allow educators to both train and undertake assessment of students. Simulators can be designed to address a number of clinical scenarios on demand and can be readily available at any time irrespective of patient throughput. According to Haycock (2011) the main potential offered by simulators is that unlike real life they can provide a number of reproducible cases for assessment. Also in the case of computer driven simulation they also have the potential to provide an objective rather than subjective evaluation of procedural skills so providing a standardized platform for assessment. Because of these reasons health educational leaders are suggesting that simulation based assessments are essential, however before a tool is widely implemented, the evidence of its usefulness and validity needs to be established, (Brydges *et al*., 2015).

Assessments within higher education can essentially be divided into two broad types, formative and summative. The former is essentially assessment for learning, intended to provide the teacher and more importantly the student with feedback which is essential for improving knowledge and skills acquisition, (Schute, 2007) and therefore can be considered as a process to actively engaging students in their own learning, (Looney, 2011). Despite this potential for assessment influencing learning the evidence of any effect is sparse, (Schuwirth & van der Vleuten, 2011); for example, Grosas *et al*., (2014) found that many students fail to acknowledge that formative testing is a learning process.

Summative assessment is often referred to as assessment of learning. They can resemble formative assessments, but are typically placed at the end of a learning cycle rather than during. However, the essential difference is in the use of the evidence derived from the assessment procedure. In this instance the assessment typically does not contribute to the students' learning, and there is no feedback into the teaching of the material, (Gardner, 2012).

More recently this dualistic view of assessments has been eroded in daily practice, and assessment discourse has shifted towards the notion that assessment is a mutually constructed construct between both the learners and assessors/teachers, (Boud & Soler, 2015). Assessments are now considered to have a multiple emphasis, being expected to fulfil several purposes regardless of whether they are predominantly summative or formative in function. This multiple emphasis on assessments consists of the following components, identified by (Boud 2010: 253-254).

- the assessment actively engages students;
- it is comprised of authentic activities;
- involving students in the design of assessments;
- integrating tasks to give an assessment that takes a holistic view of what they have been learning and know what they need to know;
- allows students to engage with model answers and practice so that they can see improvement in their work;
- working with peers to foster team work;
- giving and receiving feedback.

Trying to effectively meet all these components when designing an assessment creates tensions and compromises, and because of the importance of assessment, the task design becomes very important, (Carless, 2009). This view is supported by Ramsden (2003) and Gibbs and Simpson (2004) who both emphasise the importance of finding a suitable assessment method to ensure that the students take the correct approach to the learning tasks and have quality student engagement.

Competency based assessment is a paradigm that is common in many healthcare professions including radiography and radiotherapy. It is very much a style of outcomes-based education with the requirement for both the demonstrations and assessment of competence and exists due to the public expectation that health care professional demonstrate competence when dealing with them, (Scalese, Obeso & Issenberg, 2008). Simulation, competency and competency based assessment in relation to health care subjects is dealt with further in Chapter 3.

1.4 Problem Statement

Since the installation of VERT facilities at the researcher's host institution in 2008, the challenge of the programme team has been to look at what is offered by the new system and to integrate the VERT system into the curriculum in order to aid and enhance both education and training. Although the technology was originally given to the HEIs, the course teams need to explore its potential to improve teaching and learning. In other words, a social constructivism stance is being taken by the lecturers and students with both being given opportunities to adopt or reject the various uses that they can see for VERT regardless of the reasons for which it was supplied. One potential area that has been identified where VERT may be of use is in electron setups. Within radiotherapy, most treatment setups use predefined positions for the collimator and gantry alignment but others, such as those generally used for skin treatments, require the positioning of an applicator parallel to the surface of the patient "skin apposition". The accuracy of this type of "free" setup is determined by the expertise and skill of the radiographer, requiring very good hand eye coordination and 3D spatial ability and so requires a different set of skills to most treatment setups. These techniques are not common and are sometimes pre-calculated by planning radiographers, with the implication that learning opportunities and occasions to obtain competencies on this type of setup are limited. VERT potentially could be used in the preparation of students for a competency based assessment by allowing the students to learn and practice simulations of electron setups or alternatively as a method of competency based assessment for the "free setup" of electron fields. The stimulus for consideration of the latter not only being the relative infrequency of this type of setup, but also being able to standardise the assessment, not only in terms of standardising the setup, but also to reduce the perceived variation in marks between assessors and hospital sites which had been highlighted as an issue of concern by the students their course feedback in previous years.

VERT as an assessment tool offers the opportunity for both formative and summative assessment, however its introduction needs to be carefully considered as although it meets a number of the ideals mentioned in Section 1.3, such as engaging students and having authentic activities, it is essential to establish evidence to support its use in this manner and to involve students in the design of assessments prior to its possible implementation. This is a stance that is supported by the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

As well as looking at the use of VERT to measure competency, the study also investigated if the system was acceptable to students as a method of assessment and asked how they felt the system should be used.

1.5 Aims and Objectives

Simulation is a key tool in learning and assessment and is used frequently in many health based programmes. The overall focus of this study was to compare the use of two simulation methods to investigate firstly if similar results were obtained on both simulations and secondly to discover the students' perceptions of both methods.

In order to investigate these overarching aims the following research questions were answered. The first six aims were investigated using quantitative data and as such the enquiry was based on looking at the null hypothesis (H0) where possible. The final two aims were investigated by looking at the qualitative data.

1. To discover if Immersive tendency predicts the feeling of presence in both simulators.

$H0_1$ There is no relationship between immersive tendency and presence.

2. To assess if Presence will be higher on the LINAC simulation compared to the VERT simulation.

$H0_2$ There will be no difference in Presence scores between the two simulation methods.

3. What student characteristics moderate the presence scores?

$H0_3$ Age and gender will not affect the presence score.

4. Do the outcomes of the two simulated setups utilising different equipment agree with each other?

$H0_4$ There will be no difference in simulation setup parameters for the two simulation methods.

5. What factors moderate the simulation scores?

$H0_5$ Cohort, age, gender and immersive tendency will not affect the simulation score.

6. Do participants utilise the same cognitive process on both systems?

7. What do students think about the appropriateness of both methods as an assessment tool?

8. What do students think about the use of VERT as a method of learning?

1.6 Significance of the Study

Most of the published work considers the use of VR systems as an educational tool for the delivery of information to the user, the VE being used as a learning environment allowing users to develop and practice skills rather than as a means of measuring competence. This study undertakes the next logical step, if research indicates that we can train students using virtual reality and VERT, can we assess the students' competency via this technology as well?

1.7 Positionality and Reflexivity

Bartell & Johnson (2013) state that we cannot separate ourselves from whom we are, and that what we are has been influenced and shaped by factors such as race, gender, class as well as what society has placed upon us and our own life experiences. All researchers are positioned whether they acknowledge it or not, and their positionality will affect how they perceive their role and practice, which is what makes it so important for qualitative studies where the researcher is part of the process. Part of the ongoing debate on positionality is that of the concept of insider/outsider positionality, and the apparent advantages and disadvantages offered by each status such as insiders possessing deeper insight into field at the expense of being more biased, whereas an outsider would be more objective, possibly at the expense of lack of understanding, (Chavez, 2008).

Because of the issue of positionality it is suggested that researchers acknowledge their own positionality through reflexivity, "critical self-reflection about one's own biases, preferences reflections and preconceptions." Polit & Tatano Beck (2012: 740), so as to offer a transparency to the methods used.

The majority of the researcher's work to date has utilised quantitative methodologies and have been firmly positioned in positivism. This position predominantly stems from earlier experiences; the researcher having first studied biological sciences and worked for a period as a medical laboratory scientific officer and microbiologist, all of which was based firmly in rigour of method and standardised interpretation of results. Because of this background there is a natural tendency of the author to work with quantitative methods, presumably due to experience, familiarisation and security over any actual methodological requirement.

However, positionality goes beyond an understanding of just one's self and it must be acknowledged that the students themselves are positioned and that the researcher's position with them in not just based on the experience of researcher's past, but on other factors. As the study was undertaken in the researcher's host institution there would be "insider" positionality, which would be varied for different students. The researcher has different roles with different students, personal tutor and link lecturer to some, lecturer and programme manager to others. Depending on which role the student felt was dominant could affect how the student would perceive the researcher, affecting the researcher/subject interaction on a number of levels. Firstly it might have led to a selective bias based on expectation, which could have posed a challenge to the research process. Secondly the different relationships could negatively affect the data collection process, the students who knew the researcher better being more relaxed in the researcher's presence during data collection, so finding it easier to perform the tasks and talk more freely in the focus groups than other students so impacting on the nature of the narrative form. This familiarity and closeness to certain student groups did appear to affect the sample. Not all students volunteered for the study, but all the students who were on the researcher's clinical sites did, perhaps because they were more sympathetic to the researcher's personal position or alternatively because of the belief they may be more disadvantaged in some way if they didn't volunteer.

In all instances (both at officials' and community members' levels) the researcher tried to make their position clear during the research process. Despite this three student's directly implied that they felt that that there was an expectation in undertaking the study, rationalising during the study that the expectation of the research was to have VERT succeed as an assessment tool during the work-down after the simulation set-up. These comments came from students close to the researcher and were made outside of the data collection process. From a quantitative perspective this could be considered as an overt manifestation of bias within the study. The most likely bias that would be introduced in this situation is a type of bias known as expectancy bias or the “Hawthorne effect”. The Hawthorne effect is a known issue with any participatory observational research, (Coombs, 2003), the effect occurring when subjects adjust their behaviour simply because they are being observed, and is time dependent, the effect fading with familiarisation of the observation over time, (Walker, 2005).

The Hawthorne effect is rarely quantified as part of the research process, (Fernald *et al*. 2012) however, a number of established methods exist in order to mitigate or at least reduce the effect. Oswald, Sherratt & Smith (2014) suggest two major mitigating techniques based on existing work, firstly building a relationship and establishing a rapport with the subjects being observed. In this work this relationship and rapport was already established and it is because of this rapport that this specific element of the Hawthorne effect came to light. The second suggestion is the process of triangulation, which is possible in this instance because of the study design which allows a cross check of the findings between the two different types of data collected. In this case the results show students performing a lot poorer on VERT despite the students’ expectation and from the qualitative feedback the feeling that the VERT simulation was not as good as the LINAC based one. The likely inference from this is that the direction of difference found within the study is unaffected and correct, but that the actual degree of difference being observed might not be as large as in actuality.

Finally it is worth noting that the researcher has a very different background to the students. The researcher was qualified in radiography before many of the students were born. Also, the majority of the participants were women and there is a predominant Muslim culture amongst the students. All these differences could have affected the researcher’s reflection on the actions and behaviours of the students. Here the use of the mixed methods approach assisted to strengthen the objective/outsider positionality of the researcher during the data collection and analysis because of the existence of different data collection methods particularly in the quantitative data collection where checklists were used for the observation of the participant's setups. In the focus groups the researcher was cautious to avoid his influence in the research process and used interview guides to help ensure that both the relevant data were collected and the researcher's involvement in the construction of the data and its meanings was reduced.

1.8 Outline of the Thesis

This thesis is presented in eight chapters. In Chapter 1 the potential of VR for educational assessment is posed, specifically the use of VERT as a competency based assessment tool for electron setups in radiotherapy. The specific research questions are presented and an indication of whether the questions would be answered through analysis of quantitative or qualitative data.

Chapter 2 gives a short background introducing the reader to the profession that is radiotherapy. First the chapter looks at the development of both the profession and the education of its members. In the latter sections of this chapter the reader is given a brief overview of radiotherapy, especially that of electron treatments in order to help understand the clinical context of the study.

Chapter 3 reviews the relevant literature on virtual reality and explains what VR is. Firstly it considers the types of virtual reality and how virtual reality and other methods of simulation are being utilised in education, with particularly emphasis on their use in health care education. The chapter then returns to VR and considers the key features of the systems, fidelity, immersion and presence considering how these factors impact on learning. The final sections in this chapter consider assessment of competence and simulation, especially VR considering the validity and reliability of VR as an assessment tool.

Chapter 4 introduces the reasons for selecting the research design, a mixed methods approach, which was primarily chosen because of the ability to triangulate the results so as to give a more comprehensive answer to the question of the suitability of VERT as an assessment tool. Specifically it outlines the pilot study and how it influenced the final research design. It also details the study population, instruments selected, data collection procedures and data analysis techniques to ensure the reliability and validity of the study. The chapter concludes with a section on the positionality of the researcher within the study.

Chapter 5 presents analyses of the quantitative data collection, based on all the empirical data collected comparing the use of both simulation methods. In order not to waste data this was done firstly by comparing the first use VERT and LINAC data allowing an unpaired analysis. A second paired analysis was then undertaken on subjects who completed both simulations.

Chapter 6 focuses on presenting the data from the focus groups using thematic analysis. This section focuses on the students' perceptions of the two simulations, the use of the two simulations and their appropriateness as an assessment method considering the four identified themes, equipment; reality; the learning opportunities afforded by the VERT system; and assessment of competence.

Chapter 7 discusses the results presented in Chapters 5 and 6 relating them back to the study's aims whilst attempting to triangulate the data looking for reasons in the qualitative data to explain the quantitative data more fully.

The final chapter, Chapter 8, provides a summary of the research presented in Chapters 5 to 7 and considers both its contributions and limitations. The main finding of the study was that both simulation methods led to different measures of competency, the subjects performing better on the LINAC simulation. A number of issues were found with the VERT simulation which probably fed into identified differences in the approach to the setups in the first and final 30 seconds of the setup and their performance as measured in the quantitative part of the study. Future directions on the use of VERT as an assessment tool are also considered including recent changes adopted by the department in response to this work.

Chapter 2 - Radiotherapy and Education

2.1 Introduction

Radiography is a relatively new profession, having come into existence soon after two major discoveries in the late nineteenth century. Firstly, on November 30th 1895, Wilhelm Roentgen announced the discovery of x-rays, (Lederman, 1981) and a few months later, on the 2nd of March 1896, Henri Becquerel reported the discovery of radioactivity, (Blaufox, 1996).

The medical profession quickly realised the potential of the two new types of radiation both for diagnosis and treatment, with the first treatment reportedly being undertaken by Grubbé in January 1896, (Orton, 2013). In order to try and organise radiological work the Roentgen Society was established in England in 1897, permitting - “after vehement discussion” (Pasveer, 1989: 364) - both medical and non-medical members to join, in order to achieve a non-clinical bias. This however was opposed by some medical members who split from the Society in 1902 and founded the British Electrotherapeutical Society, which became, in 1907, a section under the Royal Society of Medicine.

The use of ionising radiation for the treatment of disease (Radiotherapy) was originally undertaken on many different types of conditions including tuberculosis, excessive sweating and ringworm, but modern use is restricted almost exclusively to malignant tumours. This restrictive use of radiation is due to both an increase in the understanding of the radiobiology of x-rays and radioactivity, and the recognition of the potential side effects of radiotherapy treatments such as tumour induction, (Kunkler, 2003).

Radiotherapy is currently one of the main treatments for most cancers (Delaney *et al*., 2005) and the demand for its use is increasing as the population becomes more elderly and the annual incidence of cancers increase (Office for National Statistics, 2012). In the UK the average annual incidence of cancer was 322,923 per annum between 2008 and 2010 (*ibid.*), a significant rise from the 2001-2003 figures. With this increase in demand for radiotherapy has come the need for more equipment, new radiotherapy centres and more staff to operate the new machines.

2.2 Professionalization and Institutionalisation

With the almost instantaneous recognition of the medical potential of x-rays, implementation of this new technology occurred simultaneously within hospitals throughout the world. The next few years saw a rapidly expanding clinical demand for x-ray treatments, Suit & Loeffler (2011) reporting 923 radiation treatments being performed at the Massachusetts General hospital in 1924 which increased to 1,220 a year later.

In the early years of radiography both the purchase and operation of x-ray equipment was unregulated, and as equipment was purchased hospitals called on nurses, doctors, porters and in some instances handymen to use the equipment, (Witz, 2004). Brecher & Brecher (1969) observed in a US context that, in 1910, radiography "was staffed primarily with younger men, most of whom were still in school or medical school when Rontgen's discovery was announced" (*ibid*.: 104). Many of these young men "had begun working with the X-rays in 1896 or 1897 as physicists, engineers, electricians, photographers" or technicians and had qualified as doctors "specifically for the purpose of qualifying as radiologists" (*ibid*.).

The process of professionalization was disrupted by the outbreak of the First World War when, "anybody, whether lay, physician or engineer, who was in possession of the apparatus performed radiological work for war purposes" (Pasveer, 1999: 366), but resumed post-war when: "established radiologists started pleading for good practical and theoretical education for X-ray workers in the medical as well as the physical sciences, complaining about the low status of radiology and the inadequate location of Rontgen departments in hospitals" (*ibid*.). Eventually new professions evolved; on the diagnostic side the radiologist and the diagnostic radiographer; and on the therapeutic side the radiotherapist and therapeutic radiographer. The radiologist and radiotherapist professions were subsumed into the medical fraternity whereas for the two new radiography professions the Society of Radiographers was formed in 1920 to give professional status to the non-medical workers in the fields of radiography.

The Society of Radiographers (SOR) was originally active in the training and certification of radiographers. In 1932 the first of a number of hospitals were inspected and officially recognised as training schools, (Society of Radiographers, 2003). The syllabus the schools followed was set by SOR and students were all examined by the SOR for qualification purposes who awarded the Diploma of the College of Radiographers (DCR) allowing state registration. However, the professional role of radiographers, particularly radiotherapy radiographers is not well understood by the public and radiotherapy itself has a low public profile, (YouGov, 2011). Sim & Radloff (2009) also suggest that radiographers have a lack of professional recognition and professional respect from other healthcare practitioners.

In 1977 the SOR became registered as a trade union and a new charitable company set up called the College of Radiographers to take over the educational and professional responsibilities (*ibid*). At about the same time the debate started about degree education for radiographers, one of the original proposals being that the degree course should be 4 years 5 months long and not only involve a degree qualification from the University, but also the DCR in order to confer state registration, (Jordan, 1995). The debate continued within radiography which together with fellow paramedical professions began to call for graduate courses. This proposal was opposed by the Department of Health and Social Security (DHSS) at the time stating that,

> "there seems only the most limited scope in radiography for the academic orientation of degree training….. it would be difficult to keep a balance between academic and vocational training in radiography…..without further lengthening the training, with no real benefit to the NHS." DHSS, (1979: 1).

Despite this opposition by the DHSS to degree status for most paramedical professions physiotherapy became a graduate entry programme, so establishing a precedent. However, even with this precedent the need for a graduate programme was still not universally accepted within the various fields of radiography and it was not until 1987 that the degree rationale was published calling on radiography schools to collaborate with Higher Education institutions. In 1990 the hospital based schools started to move into higher education establishments so creating graduate entry training programmes in radiography. Nixon (2001) remarked that, "Radiography was the last of the major health professions to move to graduate entry and managed to achieve almost total transition in a single co-ordinated move during 1990–1991" (*ibid*.: 32). With the move from a diploma based course to graduate courses the curriculum had to change and for the first time different institutions had the freedom to set their own curriculum content rather than use that set by the COR.

The world's first radiography journal was the "*Archives of Clinical Skiagraphy*," first published in May 1896, (Mould, 2011). In the following years the number of journals devoted to radiography and radiotherapy increased, but the journals were aimed primarily at oncologists, radiologists and physicists rather than radiographers. With the move to a graduate programme the profession recognised the need for its own knowledge base as the vast majority of the existing knowledge base for radiography was built on evidence produced by medical practitioners and physicists (Nixon, 2001). As part of the effort for radiography to be seen as having a professional standing a peer reviewed journal aimed at radiographers was introduced in 1995 in order for radiographers to have a vehicle to publish their own research.

With the introduction of graduate entry radiography qualifications the COR's role changed and it became a validating body for all radiography programmes in the UK. The Health and Care Professions Council (HCPC) who provide state registration for radiographers in the UK also developed an active role in course validation. Currently all validations of UK Radiography courses involve the host University, the COR and HCPC. Each of the external validating bodies (the COR and HCPC) have developed a set of competencies that must be demonstrated by the students during training and be included in the training programme. Everyone entering the profession must also be able to demonstrate the knowledge and skills stipulated in the NHS Knowledge Skills Framework (KSF) and so at least in part all radiography programmes can be said to have competency based curriculums. This form of radiography education is not unique to the UK and is also prevalent in many countries including Australia and the USA, (Leggett, 2015; Australian Institute of Radiography, 2013). In some other countries the radiotherapy radiography profession is not officially recognized and no formal education programme or registration process exists, although it is recommended that within such courses there is a need for the curriculum to define core competencies, (International Atomic Energy Agency, 2014).

2.3 Radiography Education

"Radiography education has followed the lead provided by nurse education whereby initial and post-registration provision are now located in institutions of higher education", (Castle, Holloway & Rage, 1998: 333). Prior to radiography becoming a graduate entry profession, the distinction between radiographer and radiologist as drawn by Mackay, (2003) was that, "Radiographers were educated for two or three years up to diploma level whereas radiologists were medically qualified and had several years' experience in medicine or surgery before becoming radiologists" (*ibid*.: 93).

Middlemiss (1973) noted the potential need for the development of the radiographer's role due to a potential shortage of consultant radiologists, suggesting an advanced radiologic technologist who could, "perform some of the procedures and undertake some of the responsibilities at present performed or undertaken by radiologists" (*ibid*.: 804).

This shortfall in consultants did indeed come to pass, and the COR was keen to further develop the role of the radiographer. This brought about role extension possibilities such as radiographic reporting. It also brought about a number of changes to the staffing model in the UK and the introduction of a four tier structure, which is now referred to as the Career Progression Framework (CPF). Qualifying students would enter the profession as practitioners and could progress to advanced practitioners (MSc level) who would be autonomous (DoH, 2003) and eventually consultant practitioners (PhD level) each of which are state registered titles and regulated by the HCPC. A new level of assistant practitioner was introduced below the practitioner level which would be a sub-professional level that could only work under the supervision of a practitioner.

In order to support the structure Universities had to produce programmes for assistant practitioners and as the CPF stipulated a continuum between levels they also had to produce a route from assistant practitioner to practitioner status as a means of widening participation for the profession, (DoH, 2003). Also, in order for practitioners to progress through to advanced and consultant levels post graduate modules and programmes needed to be developed up to Doctoral level to suit this purpose.

The final change that affected educational provision was the recognition that there was a need for radiographers to undertake continued professional development in order to maintain and develop throughout their career the capacity to practice safely. Until 2005 there was no requirement for radiographers to undertake any personal development post qualification and there was a general ambivalence towards Continuing Professional Development (CPD), (Henwood, Yielder & Flinton, 2004). However, in 2005 for UK radiographers it was legislated as mandatory for continuing registration with the HCPC, and a sample of registrants are checked every two years to ensure compliance. Despite this legislative change the attitude of radiographers towards CPD activity remains largely unchanged, (Henwood & Flinton, 2012). Both CPD and the CPF could also be said to be key in the professional nature of radiography as both promote the application of advanced learning and expertise to provide a service to clients, which according to Madden & Mitchell (1993) are key concepts of a profession.

Presently within England and Wales most students who become eligible for HCPC registration upon qualification pursue a three year therapeutic radiography programme leading to a BSc degree, although a small number of institutions do offer a two year post-graduate route to registration. Although the curriculum varies between the numerous institutions most programmes essentially consist of a roughly 50:50 split between academic and clinical time. The academic learning is facilitated by radiographers with higher degrees and a teaching qualification, whereas the clinical learning is facilitated by the clinical staff, most of whom have no teaching qualification. The clinical part of the curriculum at City University London as with most other institutions offering radiography programmes takes the form of a portfolio that prescribes the objectives and competencies that students need to have achieved during each year of the students' clinical education. The objectives and competencies are derived from HCPC and SCOR documentation.

2.3 The Practice of Radiotherapy

Modern radiotherapy can be classified as either brachytherapy, which is very short distance treatment, or external beam radiotherapy (EBRT) which was originally called teletherapy (treatment at a distance). The two most common forms of brachytherapy are interstitial and intracavity treatments where the radioactive sources are either placed directly into the tissue "interstitial" or into a body cavity "intracavity" (Suntharalingam, Podgorsak & Tőlli, 2005). This type of radiotherapy is limited to cancers in specific body areas and is very uncommon when compared with EBRT which most patients will receive.

EBRT is usually delivered by one of two different types of machine. Kilo-voltage units which, due to their limited penetration, tend to only be used for small superficial tumours such as skin lesions (Flinton, 2009), and the linear accelerator (LINAC) which is currently used for most treatments and is the unit utilised in this study.

The workflow for most radiotherapy treatments involves the patients undergoing four distinct stages, localisation, planning, treatment and verification. During each of these stages, although the patient will be under the care of a multidisciplinary team, the professionals with most patient contact would be the therapeutic radiographers.

Localisation is the first step in the treatment pathway and the first consideration is the patient position and whether immobilisation is needed (Griffiths & Short, 1994). The optimum patient position is important as it should allow entry and egress of the radiation beams whilst being reproducible and allowing the patient some degree of comfort. Some patients, particularly patients with head and neck tumours may also need immobilisation, and devices such as personalised thermoplastic immobilisation shells are necessary to help the patient maintain the correct position throughout their treatment (Barrett, Dobbs & Roques, 2009). The next part of the process is to find the tumour location, which is usually done via a Computed Tomography (CT) scanner. The CT scan will provide cross sectional images of the body which can provide information on the tumour size and its location in relation to the body's normal organs, reference points and the surface of the patient (*ibid*).

The CT data is then transferred via DICOM (Digital Imaging and Communications in Medicine) format to the planning department. The process at the **Planning** stage is for the tumour, the patient's surface and any organs at risk (normal tissues whose sensitivity to radiation may influence field placement) to be outlined. The operator then produces an optimised treatment plan by determining the best beam arrangement that will deliver a high dose to the tumour while delivering a dose as low as possible to the surrounding structures, particularly organs at risk. At the end of the planning process a computerised dose distribution and the associated beam parameters will be produced to be used in the next stage.

Treatment is usually undertaken on a LINAC using the patient position and beam parameters determined during the localisation and planning stages. The treatment is usually delivered over an extended time period, which is both tumour and site dependent, the dose being split into smaller daily doses called fractions. The basis of this approach lies in understanding the basic radiobiological principles called the 4 “R’s”, each of which will affect the efficacy of treatment, (Mitchell, 2013). It must be acknowledged at this point that some authors such as Steel, McMillan, & Peacock (1989) suggest that a fifth “R”, Radiosensitivity be included. Radiosensitivity acknowledges the inherent susceptibility of cells, tissues, organs to the harmful effect of ionizing radiation, however unlike the other “R’s” this R does not take place in the inter-fraction interval and is not considered further here.

Firstly, fractionating the dose allows cells to **R**epair between treatments, because tumour cells have less capacity to repair themselves compared to normal tissue. The radiation damage to the tumour cells builds up quicker leading to cell death whilst normal tissues recover (Nieder & Baumann, 2011). Cells also show a difference in sensitivity to radiation in different parts of their life cycle, and giving the dose in small fractions means that cells will be irradiated whilst in different parts of their life cycle i.e. they will **R**edistribute themselves between treatments, (McMillan, 2002), so increasing the effectiveness of the radiation.

Cells are more sensitive to radiation if well oxygenated (*ibid*). Small tumours are generally well oxygenated, but as tumours grow they tend to outgrow the blood supply leading to areas of poorly oxygenated cells, typically towards the middle of the tumour mass. With fractionated treatment, the outer oxygenated cells will be more sensitive to the radiation and they will be killed relatively quickly. This shrinks the tumour mass allowing the blood vessels to get closer to the poorly oxygenated area and so start to **R**e-oxygenate cells making the previously radio-resistant cells more sensitive to the radiation and so improving patient prognosis, so increasing tumour cell kill, (Brown, Carlson & Brenner 2014).

Whereas the first three “R’s” are all positive benefits to fractionated treatment the final “R”, **R**epopulation has a negative association. The time between the treatments allows cells to divide and repopulate, the rate of which tends to increase as cell damage and cell death increases. This is desirable for the normal tissue being irradiated, but it also means that during the treatment the tumour cells will also repopulate and this can affect the treatment outcome (Fowler, 2010). The amount of tumour repopulation therefore has to be factored into the treatment dose and overall treatment time and the effect of any delays to the treatment must also be considered, (Wyatt, Jones & Dale 2008).

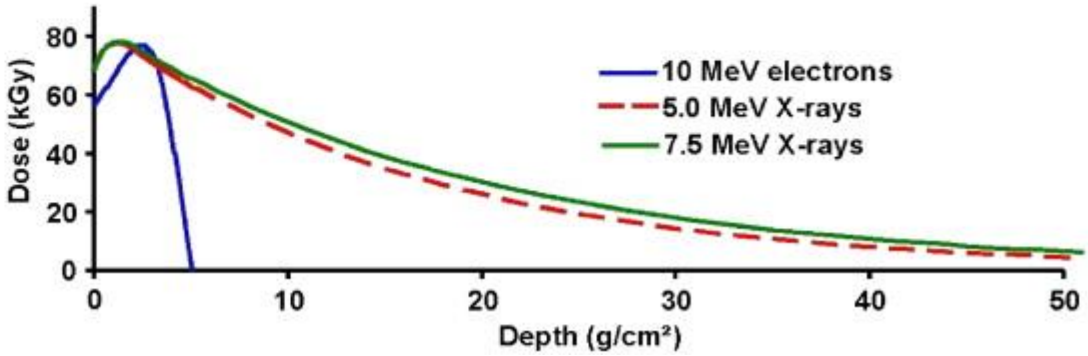
As the dose is being given over a protracted time, in some cases over 7 weeks, it is imperative that each day the fields are precisely placed in order that the dose is delivered accurately to the tumour. This needs to be checked on a frequent basis, a process called **Verification** which is the final stage of the process. Verification exists in two distinct forms, geometric - "*is the dose going to the right place?*" - and dosimetric - "*is it the right dose?*" It is currently recommended that geometric verification be undertaken for all megavoltage external beam x-ray treatments, (Royal College of Radiologists *et al*., 2008) and is usually achieved by imaging the patient with x-rays whilst in the treatment position before treatment is given.

LINACs have the potential to produce two different types of radiotherapy beam, x-rays and electrons. X-ray beams are usually used to treat deep seated tumours, whereas electron beams are used for superficial treatments as their penetration is tissue is limited as explained in the following section.

2.3.1 Electron Beams

X-ray beams deposit their energy in an approximately exponential way in tissue (see the green and red dashed line in Figure 2.1) and dose is given to the body both before the beam reaches the tumour and again as it passes through tissue beyond the tumour. This can be an issue if the tumour is situated in front of a structure that is radiosensitive as the dose to the sensitive tissue can limit the dose that can be given to the tumour. Electrons have a finite range as shown by the blue line in Figure 2.1, and so can be used to give a superficial dose whilst ensuring the dose to deeper structures beyond the tumour is very low. This makes electron beams useful in the treatment of skin tumours such as basal cell carcinoma and squamous cell carcinoma; benign skin conditions such as keloid scars; superficial metastatic lymph nodes particularly in the neck area; and as a tumour bed boost following conservative breast surgery (lumpectomy).

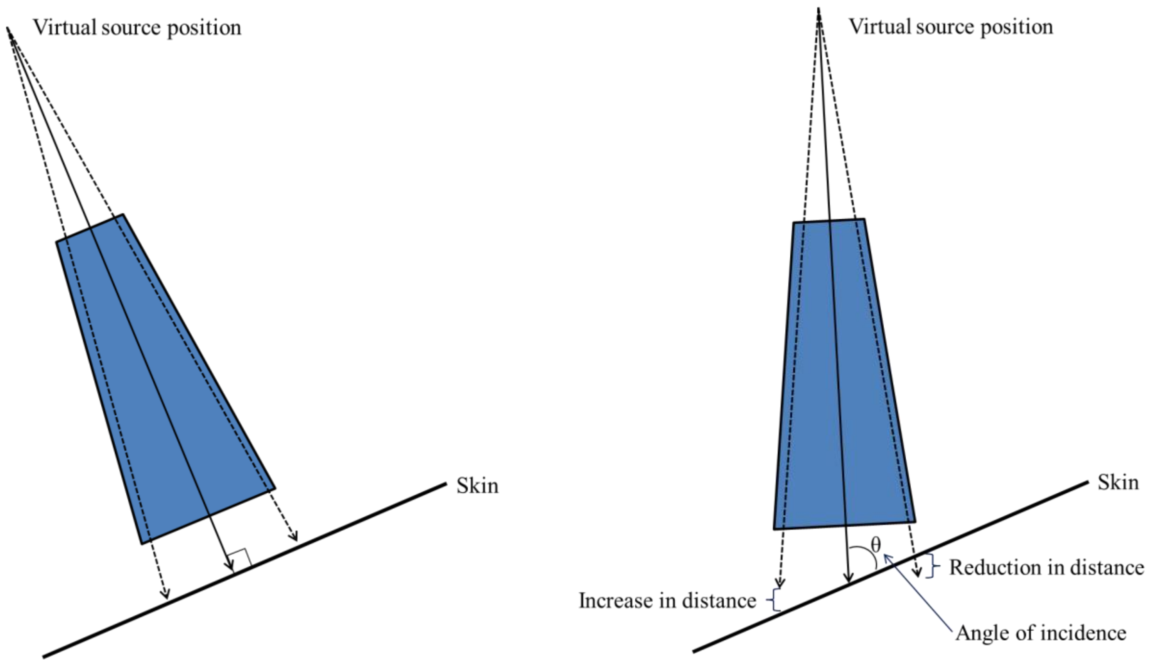
Figure 2.1 Central axis depth doses of x-ray and electron beams.



(Herer *et al.*, 2009: 532).

Electron set-ups differ from x-ray beam radiotherapy linear accelerator set-ups in a number of ways, firstly they utilise an applicator attached to the head of the unit to collimate the beam rather than the integral collimators used for photon beams, (Flinton & Miles, 2009). Secondly the set-up is not planned at set angles, but rather the beam needs to be as near perpendicular to the skin surface as possible, this is called skin apposition and an example of good skin apposition is shown in Figure 2.2a. Skin apposition is important for two reasons; any variation in skin apposition will lead to a variation in the treatment distance across the treatment field and also affects the angle of the incidence beam which is also referred to as beam obliquity; poor skin apposition is shown in Figure 2.2b, (Thwaites & McKenzie, 2007).

Figure 2.2 Electron setup and skin apposition



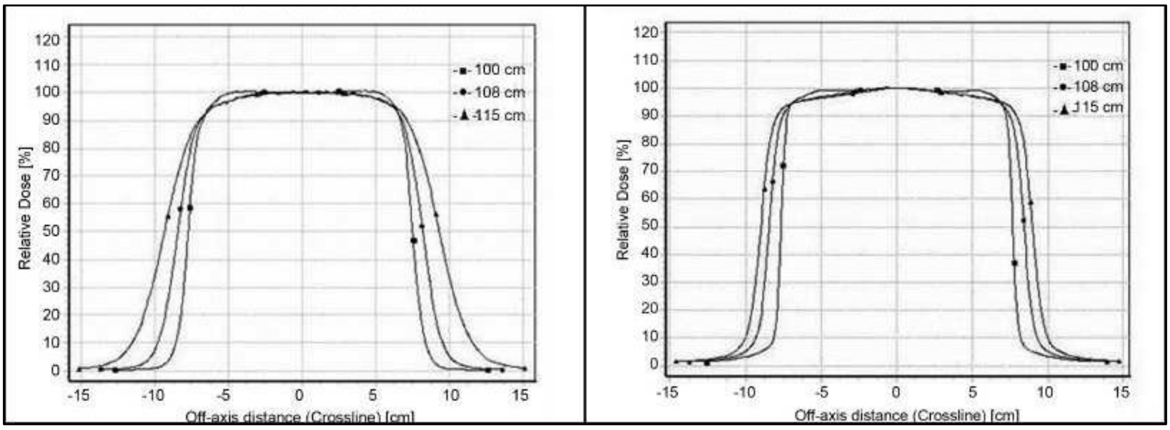
a) Good skin apposition

b) Poor skin apposition

Treatment distance: which is usually referred to as the Focus Skin Distance (FSD) is measured on the central axis of the beam, the solid central arrow in Figure 2.2a and b. Any discrepancy in the treatment distance leads to a variation in dose received by the patient as the intensity of the beam varies with the distance. For small changes in distance the effect on the dose can be calculated using the inverse square law (Strydom, Parker & Olivares, 2005), which basically states that the dose varies with the square of the distance. Larger distances give a lower dose than expected to the patient and shorter treatment distances increase the dose. The FSD is measured by radiographers using an optical distance indicator built into the LINAC which is reproduced on the VERT system. Distances to the skin off the central axis is referred to as stand-off or stand-in and is usually looked at in the four corners of the applicator. If skin apposition is good as in Figure 2.2a all distances will be equal to 100cm, and the dose to the tumour will be the same across the beam profile. In Figure 2.2b the FSD is correct but we can see stand-in on the right hand side and stand-off on the left side of the diagram leading to a dose gradient across the treatment field.

One problem with calculating the effect of changing distance on the beam is that the electron's source position is a point in space that has to be found by experimentation rather than a physical point such as the electron scattering foils and because of this is known as the "virtual source position" (Strydom, Parker & Olivares, 2005). Changes to the dose as a result of the changes in distance are relatively small, but other effects, primarily to the penumbra of the beam can also be produced. The penumbra is an area of unwanted partial dose at the edge of the beam. At increased distances the penumbra for electron beams is larger nearer the skin's surface due to the sideways scatter of electrons before reaching the surface of the patient so degrading the beam edge, (Arunkumar *et al*., 2010). At reduced treatment distances the effect is reversed and the penumbra is smaller, (Washington & Leaver, 2009). Figure 2.3 shows two beam profiles, the penumbra is the dose at the side of the beam represented by the vertical lines which can be seen is wider with increasing distances meaning that more normal tissue will be irradiated.

Figure 2.3 Penumbra variation for 2 different electron beams at three distances.



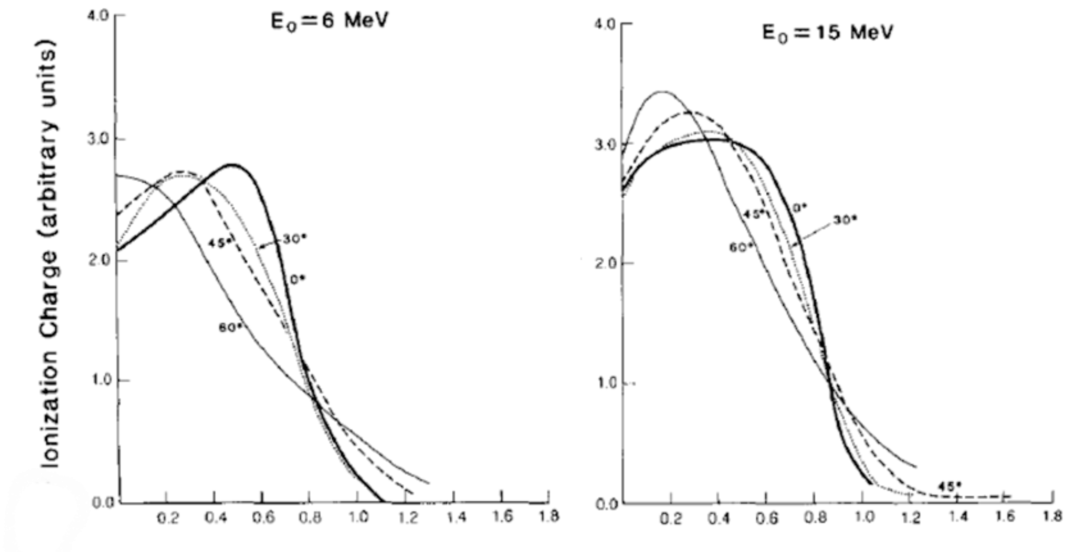
6MeV Energy beam 20MeV Energy beam

Arunkumar *et al*., (2010: 209)

Beam obliquity: The greater the obliquity of beam incidence the larger the maximum dose ($_d$max) due to increased electron fluence, but for depths beyond $_d$max dose is decreased with increasing beam obliquity and this effect is different for different electron energies, (Khan & Gibbons, 2014). Obliquity also has other effects on the beam parameters affecting the maximum distance of penetration. These effects can be seen overleaf in Figure 2.4. The highest point of the graph represents the maximum dose ($_d$max) and this changes with the different obliquities. The variation in $_d$max is more apparent at higher energy, as shown in the graph on the right. As well as the depth of

$_d$max changing the dose at depth also changes which is represented by the variation in the shape and slope of the line beyond $_d$max.

Figure 2.4 Depth dose curves for two electron energies and obliquity angles.



Adapted from Khan *et al*., (1985)

As irregular surfaces are a frequently encountered situations in radiotherapy corrections can be made to the dose at given points using the formula below, which accounts for both the change in distance and the obliquity.

$$D(SSD_{eff} + g, z) = D_0(SSD_{eff}, z)\left(\frac{SSD_{eff} + z}{SSD_{eff} + g + z}\right)^2 \times OF(\theta, z)$$

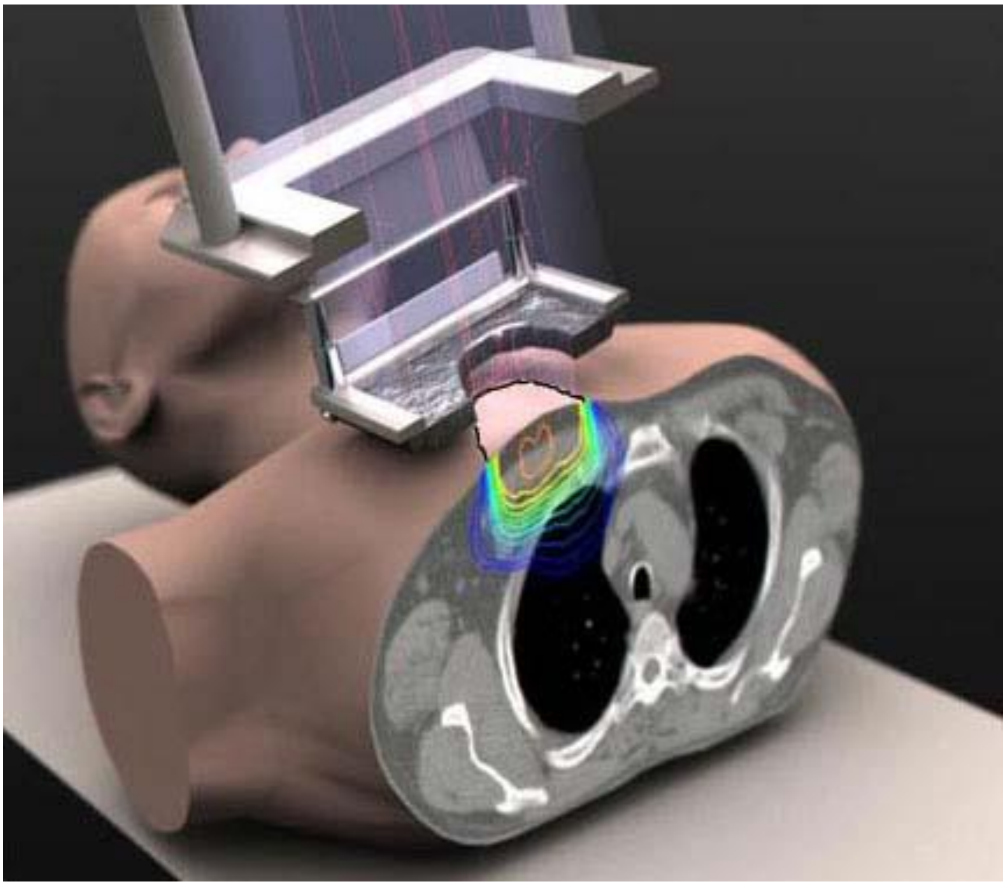
Taken from Strydom, Parker & Olivares (2005: 291)

Where: SSD $_{eff}$ is the effective source to skin distance. Note this is different from the FSD as indicated by the distance light as it is dependent on the virtual source position, which in turn is dependent on the energy and cut out used. g is the stand-off, z is the depth in the patient, θ is the obliquity angle between the tangent to the skin surface and the beam central axis. D_0 (SSD $_{eff}$, z) is the dose at depth z for a beam incident normally on a flat phantom and OF(θ, z) is a correction factor for the obliquity of the beam. This factor may be directly measured or alternatively published tables may be used.

The International Commission on Radiological Units and Measurements (ICRU) recommends that for electron beams the dose delivered to a tumour should be within 10% of the prescribed dose. The distance, skin apposition and angle of incidence have only a small effect on the dose compared to other factors such as the presence of air cavities and density changes, but these other factors are beyond the control of the radiographer whereas distance, skin apposition and obliquity are.

The final difference between x-ray treatments and electrons is that for an x-ray treatment the beam is usually given at set angles once the isocentre is set, the isocentre being a fixed point in space around which various beam parameters can rotate. For an electron treatment the area to be treated is usually drawn on the patient and the treatment field as delineated by a light beam shining down the electron applicator which is then placed over the patient so as to cover the treatment area. An example of an electron beam treatment is shown in Figure 2.5, the applicator and the end plate which are shown cut away in the figure can clearly be seen. The electron applicators come in a variety of sizes and can be tailored to a specific shape by inclusion of the end plate which is made of a low melting point alloy (LMPA). The light beam illuminates the area and is used to help the radiographers position the applicator correctly. In the example shown you can see the light beam conforming to the skin marks on the phantom and that the applicator has been angled to try and ensure that the distance to the skin's surface is equal across the field. Also superimposed on this image are the isodoses lines, lines of equal dose in the tissue; red indicating areas of high dose and blue lines low dose. The treatment shown is a high energy electron field, yet despite this you can see that the isodoses stop at a relatively short distance under the skin's surface. You can also see the penumbra represented on this image as the blue isodose lines bowing out to the side under the surface of the phantom.

Figure 2.5 Electron treatment on a phantom.



Adapted from: Northwest Radiation Oncology (2013)

The ability to fit the light beam to the skin marks at the correct treatment distance, and the amount of skin apposition can be influenced by the expertise of the treatment radiographer. All of these aspects need a good 3D perception. It is because of this and the nature of the electron set-up that this form of treatment is the focus of this study.

2.3.2 Patient care

As well as being responsible for the delivery of treatment therapy, radiographers also have a duty of care to patients. They must put patients at ease, provide patients with adequate information both before and during the treatment, (Halkett *et al*., 2012) and also recognise and give advice on the side effects experienced by the patients. Also, the majority of patients with cancer are suffering an emotional response to their diagnosis and may be depressed or anxious. Radiographers are a major support mechanism during radiotherapy, (Slevin *et al*., 1996) who see the patients on a daily basis for the duration of the patient's treatment. The radiographer must therefore not only be able to operate the equipment effectively, but also develop communication skills and a compassionate manner in order to handle these situations and give most benefit to the patient, (Martin & Hodgson, 2006).

2.4 The simulations

This section describes the two simulations, Virtual and LINAC based, explaining the similarities and differences between the two.

2.4.1 The Phantom

Both simulations used in this study represented a single patient with two electron fields, both on the chest wall. Field 1 was placed on the left anterior aspect of the chest wall and Field 2 on the right anterior aspect, see Appendix IX. Both fields were 10x7cm as currently this is the only size electron field that can be created in VERT. Field 1 was the simpler setup as the surface of the phantom is relatively uniform at this point, there being just a slight angling of the chest wall. Field two was more complicated as the surface of the phantom changes more in this area and the direction of slope changes, particularly in the supraclavicular area. There is also a further potential problem with Field 2 in that the neck and head can interfere with the applicator placement and this adds a further complexity to the setup.

Other than one simulation being real and the other virtual the phantom and position of the electron fields were identical. The position of the phantom on the couch was also identical for each simulation. VERT automatically places the simulated patient on the couch at the same point each time. This point was measured prior to the study commencing and the real phantom placed on the LINAC couch at the same point by measurement from the top of the couch prior to the start of each student's setup. The one difference about the patient between the two simulations was that on the LINAC the marks for both setups were present and could be seen by the operator at the start, whereas on VERT only one setup could be seen. Once the first setup was complete the researcher could then switch Field 1 off and switch Field 2 on. This was caused by a limitation of the VERT system which can only place one electron field on the phantom at once. To ensure parity between simulations students on the LINAC prior to commencement of the simulation were instructed to complete Field 1 first.

2.4.2 The equipment

The treatment machines used were identical for each setup, both being Varian 2100IX linear accelerators. Again the position of the start point was identical, the couch being lowered to its lowest point (-65cm), the lateral movement being set to 0 and the longitudinal position being 1.5cm. The collimators and gantry angle were both set to 0°. These positions are common start points for most patient setups. In both instances the electron applicator and lead cut out was in place before the subject started the setup. The handsets the subjects used to control each unit were identical and had identical functions allowing the subjects to control the couch and gantry movements as well as other functions such as toggling the room lights, FSD light, field light and lasers on and off. The only other physical difference in the equipment was that on VERT subjects had to wear LCS glasses in order to view the projected image in 3D. The glasses had a wire to a battery pack and sensor mount that the students clipped to a pocket. This equipment was not worn in the LINAC room.

Both simulations contained the setup parameters displayed on the LINAC centre stand. This information was repeated on monitors in the LINAC room and on the right hand side of the screen on VERT. The main difference in the equipment was that on VERT the equipment was virtual whereas in the LINAC room it was real. The couch also has controls on it that allowed movement of the couch top in the vertical, rotational, longitudinal and lateral directions both by motors and by releasing the magnetic locks and being able to move freely by hand. In VERT these controls were represented on the virtual couch, so the scene looked the same as on the real LINAC; but being virtual the controls could not be used by the operator.

Another difference between the VERT and LINAC simulation was that of the tangible nature of the equipment and phantom. In the LINAC the user could touch the phantom, which could not be done in VERT. In VERT the subject could visualise the setup differently by being able to look down the applicator to see the field setup by placing their head within the applicator, something that could not be done in real life.

2.4.3 *The room*

The room was the area the researcher had least control over in the design of each simulation. Both rooms contained lasers that were present on the walls and ceiling that could be visualised on the phantom. The virtual room however, was a pre-set generic room that could not be altered and was devoid of fixtures such as shelves, accessory equipment and cupboards which were present in the LINAC room. Also, although the VERT room was identical for each setup the LINAC room had small differences depending on the equipment needed for the patients that were currently being treated in the room. Again although there was no need for the students to do so during the setup the students could interact physically with the fixtures in the LINAC room whereas in VERT they could not. The final difference was the size of the room. The VERT room was smaller, about half the size of the LINAC room the study took place in.

Chapter 3 - Literature Review

3.1 Introduction

This chapter discusses the existing literature relevant to the study being undertaken. While the introductory chapters linked virtual reality [VR] with the constructivist learning theory focussing on students' learning, the focus of this chapter is on both simulation and VR as a tool for both learning and assessment.

According to both Gaba (2004) and Lateef (2010) simulation is not a technology, but rather a technique, which may provide a substitute for real patients, to:

> "*replace or amplify real experiences with guided experiences, often immersive in nature, that evoke or replicate substantial aspects of the real world in a fully interactive fashion*." (Gaba, 2004: i2).

This view of simulation is supported by other authors who refer to the simulator as a physical object or representation upon which a task may be performed and to simulation as the application of simulators to training or learning (Cooper & Taqueti, 2004) and that "*simulation is an educational process that can replicate clinical practices in a safe environment*" (Cant & Cooper, 2010: 3).

Simulations can be quite varied in their design, utilising a broad spectrum of tools and technologies including, case reports, mannequins, scenario discussions, and computer-based simulations. According to Sokolowski (2011), simulations can be implemented in one of three ways, Live, Virtual and Constructive. Live simulations typically involve real people and/or equipment undertaking an activity where they would for real. Virtual simulations again typically involve real people and/or equipment, but this time they perform the task in a computer-controlled setting. In constructive simulations a real person may make inputs into the simulation, but the outcomes are derived by a computer programme that determines the effect of the input on the simulated system. Another difference between these types of simulation is that for the live simulation time can be considered continuous as with the real world, whilst for the virtual simulation it can take the form of discrete steps, so allowing the user to concentrate on specific events. For constructive simulations time is not important, rather the simulation when running is driven by the correct sequencing of events to predict the outcome.

In this study both the VERT system and the LINAC setups conform to these prior definitions of simulation; both of which replicate clinical practice. The LINAC simulation would be classed as a live simulation as it simply utilises an existing treatment room and its equipment, but substitutes an anthropomorphic phantom for the patient. The VERT simulation would be classed as virtual simulation as it utilises cutting edge technology, the student interacting with a computer generated virtual reality experience.

3.2 Structure of the Literature Review

This chapter is organised into seven main sections: Constructivism; Health care simulation; Simulations and VR; Immersion and Presence in Education; Transfer of learning in simulation; competency based assessment; and assessment using simulation and VR. The first section looks at how technology and education are linked through constructivism and identifies that educational changes usually follow technological advances and takes the view of needs driving technology rather than technological imperative. The next section (3.4) looks at the current use of simulators in health care and the various ways in which they can be classified before moving onto look at the issue of simulation using virtual reality. The following section (3.5) focuses on VR looking at 3 key concepts, fidelity, immersion and presence. The latter two concepts, immersion and presence form the basis of the next section (3.6), which looks at how these concepts impact on the use of VR in education. This section first looks at the potential link between presence and learning outcomes, before considering affordance theory, which is often linked to the use of both VR and VLEs.

The final two sections consider assessments, firstly by looking at competency based assessment and then at the evidence in support of VR simulations as an assessment tool.

3.3 Constructivism

Philosophical shifts in thinking and culture may occur which can lead to re-evaluation of epistemological stances. One such instance could be said to have occurred in 1980 when David Bohm - a physicist researching quantum mechanics published a book, *Wholeness and the Implicate Order* (1980). In the book, Bohm referred to a doctrine that had been proposed thousands of years earlier by Heraclitus that everything is in a state of constant flux.

Bohm challenged the existing dominant Cartesian-Newtonian epistemological model, separating physics from philosophy and man from his environment, so compartmentalising the way we look at things, by suggesting that the universe is abstract and that the apparent separateness of subatomic particles is in fact illusory and that all things in the universe are infinitely interconnected. Bohm further argued that it followed that, "*knowledge, too, is a process, an abstraction from one of total flux*" (*ibid*, 63). This has been described as a paradigm shift, moving from a reductionism paradigm to one of holism (Davies & Gribbin, 1992).

Dabrowski's reading of Bohm's perspective (1995) is that it sustains the pedagogical challenge of how to communicate knowledge successfully in teaching by addressing schisms such as thought/action, subject/object, and knowledge/reality. It therefore informs critical constructivist perspectives on education. This is an idea that is supported by Kincheloe & Steinberg (1993) who contend that continuous etymological questioning allows both teachers and students to change and expand their roles, moving from transmitters of knowledge to researchers, so contributing to the discernment of the deeper hidden patterns of the implicate order as proposed by Bohm.

Moore (1995) states that the development of many models of learning have historically coincided with technological developments and recently technological advance has been so rapid that Cox (2013) contends teachers and researchers have struggled to keep up with what contemporary technologies offer, how they can be utilised and how they might affect the ways in which students learn. The increase in the use of computer related technologies in teaching has coincided with the growth of constructivism as an epistemological commitment, with the introduction of each complementing the other (Nanjappa & Grant, 2003). According to Lainema (2003), constructivism was originally based on the belief that technology based learning was more effective than classroom based learning. Although this concept may no longer hold true, and there are a number of different beliefs within the broad definition of constructivism, it is still felt that technology and constructivism are still integrally linked (Gilakjani, Leong & Ismail, 2013). This synergy between technology and educational methods is important, as Maddux, Johnson & Willis

(2001) explain, for the use of computers in teaching and learning should make available better ways of education that make learning tasks more authentic rather than an alternative method eliciting the same procedures and behaviours.

Whilst the constructivist view cannot be regarded as “definitive”, it is currently accepted as having the greatest relevance and the most implications for educational practice, policies and models, (Brown, 2005). The early constructivist movement in education was largely shaped by Dewey and Piaget and, although a number of different divisions of constructivism now exist, central to the idea of educational constructivism are the principles that cognitive constructions develop through action that is individual to the learner. The theory hypothesises that individuals will try and “construct” their own meaning from an experience and therefore, according to Swan (2005), all learning is linked to a student’s experience and the contexts of experience. Constructivism has its focus on the student and their learning rather than the communication of information and how students learn, so much so that Thompson (2000) suggests that constructivism is not a learning theory but rather a description/model of how people learn. In other words, for a constructivist, all learning will happen in a constructivist way regardless of method and therefore constructivists cannot prescribe ways to teach, (Proulx, 2006).

This chapter so far has outlined the perspective that technology such as VERT and VR systems in general being technology based are potentially linked with a constructivist view of learning, whereas radiography education in Chapter 2 was said to be competency based rather than constructivist. This apparent incongruity can be reconciled when you consider within competency based education it is essential that the learner develops and is able to demonstrate the required competencies. However, the teaching and learning strategies that facilitate this must ensure student engagement in all aspects of acquiring the knowledge, skills and professional behaviours, (de Kraker, Lansu & van Dam-Mieras, 2007). Competency is essential, but it is only part of the journey; students must be able to understand the meaning behind their actions and the constructivist learning theory stresses the learner’s active construction of meaning in context, so the content needs to be realistic and engaging, (Jordan, Carlile & Stack, 2008) which simulation based training can offer. This link between the two theories is perhaps best summarised in the following statement:-

> "*Constructivism is usually the dominant paradigm in competence-based learning, and learning environments that stimulate active, contextual construction of knowledge and understanding and active acquisition of competences…"* de Kraker, Lansu & van Dam-Mieras (2007: 109).

As the focus of this review is on the use of one such technology, VR simulation, the next section considers the use of simulation in health care settings and the types of simulators that can be found.

3.4 Health Care Simulation

Simulation is a technique for both practice and learning that is used by many different disciplines and trainees, but is widely used within health care education and training. Owen (2012) noted the long and eminent role played by simulation:

> "*Centuries and even millennia ago, pioneers of simulation understood how to use simulation to develop competence and confidence in students and trainee health professionals*" (*ibid*.: 102).

Pilcher & Bedford (2011) would argue that simulation began with both role playing and the use of standardised patients (actors) while Aggarwal *et al*., (2010) have argued that simulation-based training began when life-like mannequins were used as an educational tool. Contemporary simulation, Aggarwal *et al.* (2010) contended, began in the late 1960s and early 1970s:

> "*…. when researchers from the University of Miami developed Harvey, the Cardiology Patient Simulator, a hybrid between a sophisticated task trainer and computer-enhanced manikin simulator.*" (*ibid*.: i35).

Although used widely, the use of simulation is perhaps most firmly established in two key areas of training, cardiac life support and laparoscopic surgery, (Okuda *et al*., 2009). The widespread use of simulations in health related education is related to both humanistic and educational drivers, (Bearman, Nestel & Andreatta, 2013). Humanistic drivers include such considerations as the ethical imperative and patient empowerment.

Educational drivers become imperative with an increase in students undertaking the programmes, facilitation of a systematic approach, learner centred education and rehearsal of infrequent events. Nestel *et al.,* (2011) provided a list of reasons for the prevalence of simulation in health care education and training (see Table 3.1).

Table 3.1 Drivers for uptake of simulated patient-based education.

- Raised profile of patient perspectives and patient empowerment
- Ethical imperative of causing no harm to patients
- Implementation of working time directives
- Prominence of the patient safety movement
- Increased numbers of medical and health professional students
- Reduced hospital stays for patients
- Growing evidence of simulation as an effective educational method
- Growing evidence that effective health professional/patient communication is key to patient and clinician (learner) satisfaction and reduces litigation
- Development of national assessments
- Facilitates a systematic approach to curriculum activities
- Development of 'professional' competencies
- Carefully constructed simulations
    - Assure students have direct/indirect exposure
    - Allow for adjustment in the level of challenge
    - Identify boundaries of competence
    - Provide access to technical, communication and other professional skills essential for safe clinical practice
    - Enable rehearsal of infrequently occurring events
    - Assure the development of reflective practice (video, debriefing)

(Nestel *et al.*, 2011: 2)

In radiography education one of the most important considerations is that of the ethical imperative not only for the safety of the patient as mentioned above, but also that of the radiographer, simulation allowing the students to practice safely untill mastery of the procedure or skill.

At most UK institutions and in America simulation is an acceptable method of learning especially in areas such as cardiac life support, a common use as mentioned earlier in this section. It is also used for more radiographic specific areas such as patient positioning, safe equipment use, radiation safety and image evaluation. It is far less commonly used for demonstrating competency although the American Registry of Radiological Technologists (ARRT) states that if there are legal constraints or suitable patients cannot be found a simulated measurement of competency can be used if the simulation is undertaken in similar circumstances to that of the clinical procedure, and utilises the same cognitive, psychomotor and affective skills as the real event. Also, that the programme director is confident that the skills being measured will transfer to the clinical setting (ARRT, 2012). However, for all the mandatory patient treatment competencies they stipulate that these must be demonstrated on real patients (ARRT, 2013).

Over the past decade the interest in the use of simulation in many specialities within medical and para-medical professions has increased, partly as a result of the increased emphasis on the need for students to have developed some clinical skills prior to clinical practice, (Cooper *et al*., 2012). One issue frequently raised about simulation is the high cost associated with their use, both in terms of the initial purchase cost and also the recurring cost of maintenance, as well as the high staff costs due to the low staff to student ratios, (Ogden *et al*., 2007; Datta, Upadhyay & Jaideep, 2012).

3.4.1 Classification of Educational Simulators

A number of different methods exist for classifying simulators; some authors classifying simulations based on what is to be learnt. This was the approach taken by De Jong & van Joolingen (1998) who divided computer simulations into two types: simulations containing conceptual models which are focussed on learning concepts and facts and those based on operational models which concentrate on the learning the sequences of both cognitive and non-cognitive operations. Alessi & Trollip (2001) took a similar approach, the classification system being based on two types of simulation, one that teaches the user "about something," and one that teaches the user "how to do something", which are very similar to the categories identified by De Jong & van Joolingen (1998) who looked at them using the main discovery learning processes. However, Alessi & Trollip (2001) break each of the two main categories down into a further two categories. The simulations that teach the user "about something," being further divided into Physical simulations and Process simulations, the difference being that in the latter case the user can repeat the scenario using different parameters to see the effect on the outcome. The simulations that teach "how to do something" are divided into Procedural simulations where a sequence of tasks is learnt to accomplish a goal, and Situational simulations. This last form of simulation is perhaps the most difficult to produce as it models the behaviours and attitudes of people and/or organizations and therefore also has a higher expense associated with it.

Gredler (2004) offers two broad classifications of simulations: experiential and symbolic simulations. Experiential simulations are where learners interact with real-world scenarios and experience the role being simulated, i.e. the learner becomes one of the functional components of the simulation and in order to feel the experience it is essential that the simulation has a high-fidelity, *i.e.* has a high degree of visual and haptic realism, (Ioanna *et al*., 2014); this category includes simulation such as pilot simulators. Two key defining characteristics of symbolic simulations are that the students function as a researcher/investigator who is allowed to test their conceptual model within the simulation and as such there are consequences for their actions.

Alinier (2007) takes a different view of classification, looking not at the educational point of the simulation but rather at the technology being utilised during the process. She proposes 6 distinct types of simulation, which are shown below in Table 3.2.

Table 3.2 Typology of simulation methodologies.

Simulation Level	Simulation Technique
0	Written simulations includes pen and paper simulations or ‘‘Patient Management Problems’’ and latent images.
1	3-D models which can be a basic mannequin, low fidelity simulation models, or part-task simulators.
2	Screen-based simulators Computer simulation, Simulation software, videos, DVDs, or Virtual Reality (VR) and surgical simulators.
3	Standardized patients. Real or simulated patients (trained actors), role play.
4	Intermediate fidelity patient simulators Computer controlled, programmable full body size patient simulators not fully interactive.
5	Interactive patient simulators or Computer controlled model driven patient simulators, also known as high fidelity simulation platforms.

Adapted from Alinier (2007): e245.

Classifications of simulation based on the technology as above or the level of immersion, the perception of being present in the virtual world are far more common in computer based VR simulations.

According to Strangman, Hall & Meyer (2003) there are two main types of virtual reality environment.

- Desktop virtual reality, which utilise ordinary computer screens and include computer games and virtual tours that use computer graphics to simulate the real world.

- Immersive virtual reality environments that use room-size screens and use specialised equipment to allow interaction with the environment such as stereoscopic head-mounted displays and haptic gloves.

In a review of literature undertaken by Costello (1997) it was recognised that most VR systems fall into three main categories that can be ranked by the sense of immersion, or degree of presence it provides.

The most basic form of VR recognised by Costello "Non-immersive (Desktop) Systems" is the same as that identified by Strangman, Hall & Meyer. The second form of VR are the Semi-Immersive Projection Systems which comprise of a computer generating high quality graphics coupled to a large screen projection system. The system may be single screen or contain multiple screens to increase the degree of immersion. The system can be either 2 or 3 dimensional. One of the most common ways of obtaining a 3D image is the use of liquid crystal shutter (LCS) glasses. These consist of a separate lens over each eye the LCS screen blocking and unblocking the view alternatively for each eye so that each eye sees a slightly different view of the projection. This switching between images occurs so rapidly it is undetectable by the user, the brain then decodes the two disparate images and fuses them together to see one constant 3D image, (Woods, 2005). The VERT system utilised in this study would fall into this category of VR system as it was a dual back projection system with students wearing LCS glasses. The final category recognised by Costello (*ibid*.) was the fully immersive head-mounted display systems.

Thurman & Mattoon (1994) developed the idea further and rather than defining different types of virtual reality, they proposed a model for differentiating between different types of virtual reality based on three different dimensions; verity, integration and interface, that combine to form a three-dimensional classification system for virtual realities. This model has the flexibility to compare different virtual realities that don't neatly fit into predefined types of Virtual Reality such as that proposed by Jacobson (1993).

3.5 Simulation and VR

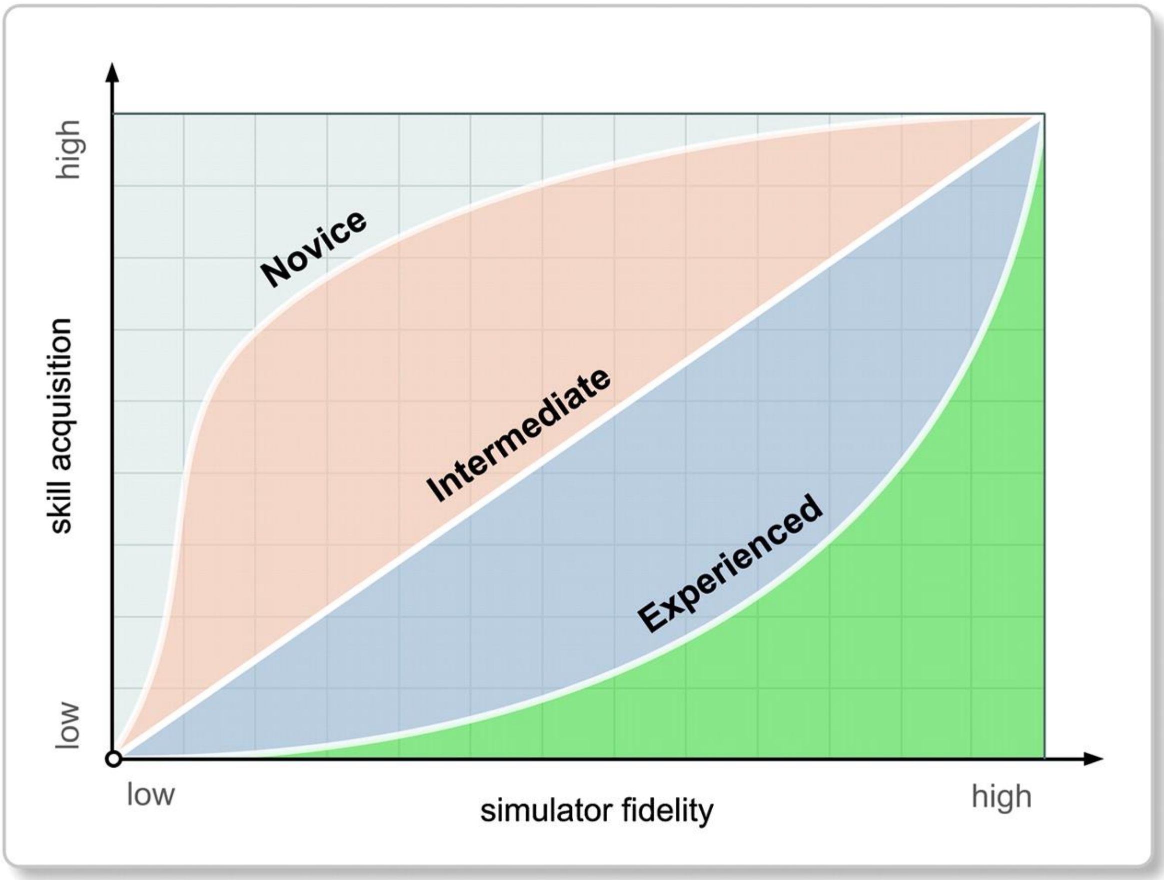
When discussing simulation and prior to looking at how VR impacts on education it is important to understand a number of concepts of VR such as fidelity, immersion and presence. These 3 concepts are thought to be key features of VR and are discussed further in the next 3 sections.

3.5.1 Fidelity

When looking at simulations, fidelity mentioned in a number of classification systems in the previous section appears to be an important issue in certain instances. Fidelity is the degree to which a simulation is loyal to that which it simulates, *i.e*. how close to reality it is. It cannot be an exact duplicate of real life or it would be the real thing and not a simulation, but it has to have some degree of reality in order to engage the user. Fidelity itself is not a single concept but is multidimensional, containing a number of different parameters such as environmental fidelity, task fidelity, physical fidelity, cosmetic fidelity, psychological fidelity and cognitive fidelity, (Liu, Macchiarella & Vincenzi, 2009; Aggarwal *et al*. 2010). Alessi (2000) suggests that low fidelity simulations are best used early on in training when looking at isolated skills, whereas high-fidelity simulation is best for advanced situations where the user is more experienced in the real world situation and also when trying undertake assessment. This view is supported by Aggarwal (2010) who consider novices gain most from low fidelity simulators compared to other level learners, but that they have a decreasing return from higher fidelity simulators. Fidelity does not necessarily reflect the level of technology being used; some proponents contend that by deliberately making some aspects of reality more simplistic, learners are better able to both analyse and understand the essential principles of a procedure and also gain the ability to perform better during the procedure, (Myers, 2012; Garson, 2009; Maier & Grossler, 2000).

Experienced learners on the other hand show an exponential increase in skills acquisition return as fidelity increase, experienced learners gaining most from higher fidelity simulations as shown below in Figure 3.1. When looking at the different types of fidelity, psychological or functional fidelity should be the most important consideration as this is the degree to which the skill or skills in the real task are mirrored by the simulated task, (Maran & Glavin, 2003). The authors (*ibid*) also mention that high fidelity simulators have been used effectively to assess technical skills, quoting studies by Chopra *et al*., (1994) and Morgan, Cleave-Hogg & Guest (2001), although this claim should be interpreted in a limited way as neither study looked at comparing performance on the simulator with identical real life scenarios, but rather at increased performance as measured on the simulator after training on the simulator; and the reliability of scoring competence on a simulator respectively. This issue is looked at later in section 3.5.

Figure 3.1 Relationship between level of experience and simulator fidelity.



Taken from Aggarwal *et al*., (2010): i36

3.5.2 Immersion

Immersion and presence are terms that some authors (McMahan, 2003) use interchangeably, whilst others view them as separate entities; immersion being but one part of the construct of presence. She also suggests that "*immersion has become an excessively vague, all-inclusive concept*" (McMahan 2003; 67). Some authors consider immersion to be psychological in nature; however the more prevalent understanding of immersion is that it is a physiological concept.

Witmer & Singer (1998) refer to immersion as a psychological state characterised by the degree to which an individual feels absorbed by or engrossed in a particular experience. Other researchers look at it from a different standpoint and have suggested that immersion is more likely to be a product of the technology itself, the technology facilitating the production of multimodal sensory "input" to the user (Slater & Wilbur, 1997; Draper, Kaber & Usher, 1998; Bystrom, Barfield, & Hendrix, 1999). Slater (1999) in particular is critical of Witmer and Singer's approach, proposing that their perspective of immersion goes against the broadly accepted concept that immersion refers to both the objective properties and physical properties of the virtual environment, whereas presence is concerned with the subjective aspects of the experience. This view of immersion could be considered as overly simplified as it suggests that technology and technology alone influences immersion, and there is therefore the assumption that different types of technology always give different levels of immersion when compared to other types of technology. This takes us away from the commonly accepted general use of the term "immersion" which may be more of a product of various factors of an experience such as expectation, plot and engagement. So a book might be more immersive than a 3D movie and vice versa dependant on these factors, Grimshaw, Charlton & Jagger (2011) finding that in some instances the level of realism can, at times, detract from an immersive experience rather than add to it, (Fogg, 2002).

Immersion in Virtual Reality may therefore not be generalizable to other media such as immersion in games and other pursuits like reading, yet this leads to further confusion as virtual reality is merely a technology that can be used for various goals such as education, assessment, rehabilitation, treatment and gaming.

In order to overcome some of these issues raised Sherman & Craig (2003) recognise two different types of immersion in virtual reality, “mental immersion” and “physical immersion”. They describe mental immersion as being deeply engaged to the point that the user suspends their disbelief in what they are experiencing. Confusingly they also refer to this as also as "having ‘a sense of presence within an environment” (p: 9). Physical immersion is described as "bodily entering into a medium” (p: 9) by stimulating the body’s senses through the use of technology.

3.5.3 Presence

Presence is often referred to using a variety of terms, common ones being telepresence, virtual presence and mediated presence. The term “telepresence” was first proposed by Gunkel (2003) in relation to the use of high-quality sensory feedback from instruments that would overcome issues in using robotic arms in industries such as nuclear power plants so that operators could feel and work with tools as though they were using their own hands. The use of the term was later refined and it is now commonly associated with the feeling of being transferred to another location, (McLellan, 2004), which is similar to the common use of the term “presence” in literature which is, “the sense of being there” i.e. in the virtual world as opposed to the real world. As such presence is about perception, a purely psychological construct and is independent of the characteristics of the technology, (Min Lee, 2004).

Presence is a multidimensional construct which depending on the type of virtual reality can include both social presence and co-presence. Social presence is defined as the sensation of interacting with other forms of intelligent agents in the VE, (Mestre, 2006) whereas co-presence is the awareness of other individuals within the virtual environment and their awareness of you, (Bulu, 2012). Co-presence can be distinguished from social presence in that social presence relates to the quality of the medium and its ability to allow communication, whereas co-presence addresses the person’s perception of others and the ensuing psychological interactions of the individuals, (Nowak & Biocca, 2003).

Presence is considered a fundamental aspect of virtual reality, (Sas & O’Hare, 2003), Riva & Mantovani (2014) considering it essential in that presence firstly “locates” the

Self in an external virtual space and once present in that space the ability to act in it. Secondly, presence provides feedback about the activity which allows the person to perceive change and act accordingly.

Witmer & Singer (1998) suggest that there are a number of key factors to users experiencing presence; selective attention, involvement, immersive response and personal characteristics of the user. The authors like others in VR research consider immersion to be a key element in presence and coined the term “immersive tendency” to summarise the characteristics of the user that influence the feeling of presence.

Other factors identified by various authors that may contribute to the user’s experience of presence are their cognitive style, the way people perceive, think, solve problems, learn, and relate to others, (Kozhevnikov, 2007) and personality, (Lombard & Ditton 1997; Alsina-Jurnet & Gutiérrez-Maldonado (2010). For example Alsina-Jurnet, Beciu & Maldonado (2005) reported that users who were introverted, or had a high degree of test anxiety, or a high spatial intelligence tended to experience a higher degree of presence. When investigating cognitive style differences in VR use Kroutter (2010) found that users with different cognitive styles (field dependents and field independents) interacted with the VE in different ways, field dependents took longer due to them being more susceptible to distraction and disorientation and therefore experienced greater navigational difficulties. However, both groups learned effectively from the VE so leading to the conclusion that differences in cognitive style have more influence on the learning process than on the learning outcome.

More recently Floridi (2005) has challenged the established models of presence, proposing a new model of presence as “successful observation”. He argues that the existing models refuse to take account of the complex dynamics between presence and absence and that rather than being a perception experienced by the user, a user is present in an environment if they are having an observable effect on that environment. In order to make the point he looks at a doctor remotely operating on a patient, arguing that regardless of the doctor perception/or lack of being present they must be present by nature of their action.

3.6 Immersion and Presence in Education

Achieving a sense of presence is considered to be a key factor for effective learning in a virtual environment, (Kroutter, 2010). This view is supported by Persky *et al*., (2009: 263), who states that presence allows for a, "*focussed and naturalistic interaction with educational materials and activities*." Picking up on this Bulu (2012) in a study looking at 3D virtual worlds found a positive correlation between presence and satisfaction. This finding was supported by Safaei & Shafieiyoun (2013) who looked at the virtual world Second Life and found that students who perceived a high feeling of social presence also thought that they learned more within the virtual environment. Caution must however be used in looking at the results of these studies as there are a number of different types of VR systems and types of presence. Also, not all studies have found a relationship between presence and educational benefit; Persky *et al*., (2009) compared the influence of presence between two types of immersive VLEs using two different teaching techniques, active versus didactic on both comprehension and engagement-related outcomes. They found that an active VLE promoted greater presence, but could find no relationship between presence and learning comprehension outcomes within either virtual environment. They did however find presence was related to information engagement only within the didactic immersive VLE and so believe presence does not operate uniformly across VLEs; rather, its effects are dependent on environmental features and so greater presence may not always relate to improved learning outcomes.

One common theory linked with use of VR and VLEs particularly in education circles is that of affordance. As can be seen from the discussion so far the virtual world the user is immersed in cannot be real, and the user's perception of the VE and objects within this environment are altered compared to reality. Affordance theory was first proposed by Gibson in the fifties and is based on the concept of direct perception of objects.

3.6.1 Affordance theory and VR

Affordance theory was first proposed by Gibson as an ecological alternative to the existing cognitive approaches. Affordance theory looks at how humans directly perceive objects in their environment. An affordance cuts across the dichotomy of the need to be either subjective or objective property, as an affordance can be considered equally to be neither or both, being a fact of both the environment and behaviour, (Gibson, 1986).

Affordances look at the relationships between an object, its context, possible actions and the effects of the actions. It offers an alternative to both the realist perspective that believes that the world exists independently of the perceiver and the radical nominalist position that reality is based only in our minds and properties only exist if they are perceived, (Hammond, 2010).

According to (Chemero, 2003) in the ensuing years the theory has been evaluated and there appears to be universal agreement between researchers that an affordance is a person relative property of the environment. For a given object the affordance is always there and has always been there, but it needs to be perceived by the agent to be realised, it is an opportunity for action. For example a small rock may be perceived by a crab as a shelter, lichen a medium for growth, whereas a human might use it as a hammer, weapon, paperweight, part of wall, decoration, counterweight etc. The rock has a number of potential uses (affordances) which are always there, but unless recognised by the user they cannot be actioned, once recognised then the user can utilise their potential. Also the affordances offered by an object are different, depending on the user's expertise and there is a general belief that new affordances become apparent to the user as they continue to use and interact with the tool, (Still, 2009).

Affordances offered by technologies; of which VR is one allows either embodiment or constrain by presenting the user with a situation and possible actions that allow the user to interact and present different behaviour patterns in relevant ways. Extending this line of thought further, this would indicate that we can not only consider immersive VR as an educational tool, but also as a place. Originally Gibson considered affordances as "action possibilities", but more recently this has been expanded. Discussing only action, what a user can do in the world is quite a limiting construct of a technology's affordance and there has been a shift to the transactional approach to affordances which focuses on the technology as a place looking not only at what the user acts upon, but also what the user receives back from the world.

Most affordances offered by technology are functional in that the potential action and reaction are performed to support some purpose or goal, (Markus & Silver 2008). Other forms of affordances may exist in each different situation. For example within this study the goal was to look at the measurement of competency, however the students could recognise the moral affordance offered by the simulations of not harming the patient. This was most apparent in VERT where students could save time in the setup by moving the applicator through the patient rather than dropping the couch to ensure patient safety before moving the applicator.

A number of authors have tried to classify affordances. Kirschner *et al*., (2004) present a framework that defines both technological affordances and social affordances:

- Educational affordances: characteristics of an environment and the user interaction with the environment that have the potential to enhance the learning potential of the user:
- Social affordances: aspects of the environment or object that provide and promote the learner's social interaction.

Hartson (2003) however used Norman's (1986) stages of action model to identify four different complementary affordances; cognitive affordance, physical affordance, perceptual affordance, and functional affordance. A cognitive affordance is a feature of an object that helps aid or support thinking whilst a physical affordance is a design feature that enables the user to do something. Hartson acknowledges that with the latter concept physical affordance can only occur with physical objects and so includes active interface objects as physical objects as they can be interacted with to initialise an action. Perceptual affordance refers to the function an object clearly offers the user within the environment, whilst functional affordance relate to a system's reactions or outcome to user actions. The author also proposes the concept of sensory affordance that feeds into the four main affordances, sensory affordance being a design feature that helps or enables the user to sense something.

Gaver (1991) looked at affordances from the interface design perspective and the possibilities they offer the user identifying three categories of affordance; perceptible, false and hidden. This simple classification system identifies perceptual affordance, a positive outcome reached by the user identifying correctly a direct link between an object and an action and two alternative affordances, false and hidden that lead to mistakes. Examples of false affordances are objects that look like buttons, but are not or the use of blue underlined text on a web page that is not a link. Hidden affordances arise when there is no information available for an existing affordance and its existence must be inferred from other evidence. Both false and hidden affordances have the ability to alter the user's sense of control within the simulation as operations may no longer be intuitive; this in turn will have an effect on presence, which itself is linked to the user's sense of control and intuitiveness of interaction within the system, (Riva & Mantovani, 2012).

Within VR the user enters a virtual simulation and as such there are no real affordances, but rather virtual affordances which Blom (2010), argues under certain circumstances are different to real affordances. Defending this statement he cites two papers, firstly de Kort *et al*., (2003) who compared cognitive mapping of a virtual environment to the real world equivalent. Although not the main thrust of the study the authors found that the users' perceptions of the environment in regards to what activities were afforded by the two locations differed, the real world being more frequently associated with social activities, while the virtual world was tied more often to formal activities. The second study by Albrecht, Blom & Beckhaus (2009) focussed directly on perceived affordances of objects in two identical settings one virtual, one real. Although virtual and real objects mostly had the same perceived affordances there was an increased tendency with virtual objects in terms of playfulness and destructive tendencies. Two factors, gender and gaming experience were found that influenced this finding.

These findings would appear to indicate that perception and action for some objects are different in the virtual world compared to the real world and dependent on a number of factors that do not affect affordance in the real world. Blom (*ibid*) also adds that the consideration of virtual affordances is important as the users in most instances intuitively perceive the affordances of an object to be the same as the real object and therefore the expectations of objects has to be met, if the expected affordance, such as being able to move an object is missing then the users experience is going to suffer.

3.7 Simulation and Transfer of Learning

The ultimate goal of any training programme is that the learning can be transferred to real world situations. A number of recent meta-analyses have been performed looking at transfer of learning in different medical specialities; Ma, *et al*., (2011) reviewed the evidence for catheterisation simulators; Mundell, *et al*., (2013), resuscitation; and McGaghie, *et al*., (2011), all medical simulation if compared with deliberate practice.

In general, all the meta analyses support improved skills acquisition with simulation based training. One issue with much of the evidence reviewed by Mundell, *et al*., (2013), is that in 78 of the 114 studies, they compared the effect of simulation with no intervention. Effectively this meant that the majority of studies that were included in the analysis had no control group, the design best being described as a case series (an uncontrolled longitudinal study). Although models of ‘hierarchies of evidence’ vary and remain contested it is usual for this type of evidence to be in either the penultimate level of evidence or as suggested by Bagshaw & Bellomo (2008) the lowest level. The rationale for this is that case series are disposed to a range of biases, including sampling; selection; detection; reporting; observer and measurement. The design is considered so weak that some authors (Rosenfeld, 2010) suggest that case series should not be submitted/published under original research, but rather as a short scientific communication which also could heighten publication bias associated with this form of research. This stance that the design is very weak is supported by Dalziel, *et al*. (2005) who state that traditionally they are excluded from most systematic reviews. From a medical and para-medical perspective case series are only used for innovative treatment or first evidence and are followed by clinical trials to substantiate the observations where they can have a significant impact on subsequent literature, (Albrecht, Meves & Bigby, 2005), however, within education randomisation and blinding are often impossible to attain and the paradigm educational researchers work from is often not the same as those in health care. In fact Hitchcock & Hughes (1995), state that this form of research is appropriate in an educational setting as its principle rationale is to reproduce social action in the student’s usual setting. Having said this any education intervention is likely to illicit a change in the population compared to doing nothing, so although a large effect was observed the result must be treated with some scepticism. This methodological design was also included in the meta-analysis undertaken by Ma *et al*., (2011) which made up 9 of the 20 studies and McGaghie *et al*., (2011) which included 4 such studies in a sample size of 14.

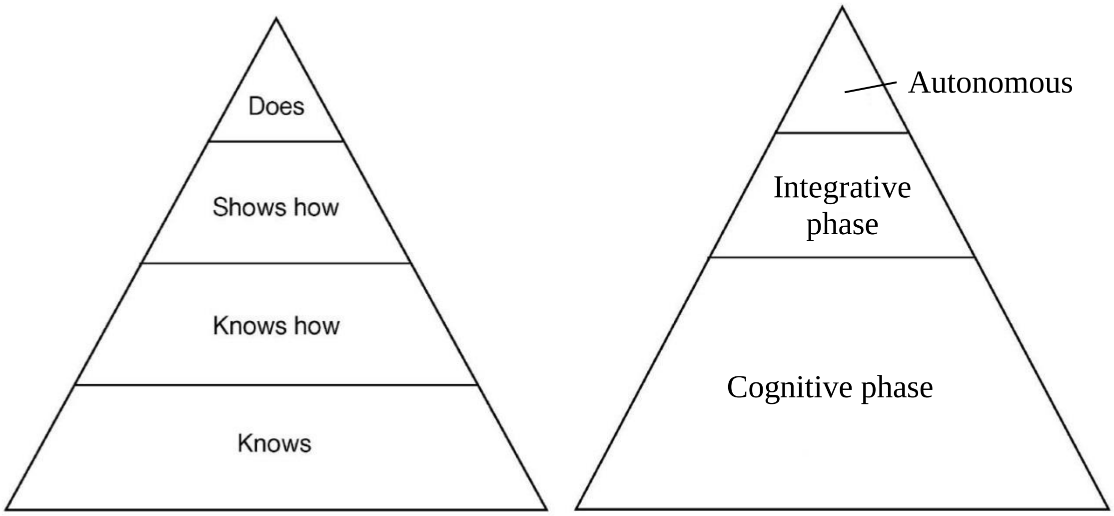
Higher levels of evidence however do exist of the efficiency of simulation as a learning tool compared to other teaching methods; Mundell *et al*., (2013) for instance includes 21 studies that compared simulation with other instructional modalities, although no mention is made of randomisation. They discovered a pooled effect size which showed a moderate to large effect favouring simulation for satisfaction, but only small effect sizes for knowledge, process skills and time outcomes. Ma *et al*. (2011) also report on 3 two group studies all of which were non-randomised which showed that simulator based training was associated with the reduction of two negative indices, the number of needles needing to be inserted into the patient and risk of pneumothorax although there was no reduction in infection risk. McGaghie, *et al*., (2011) had the highest percentage of studies that utilised a control group which was present in 10 of the 14 studies included in the meta-analysis and they describe their findings as clear and unequivocal; that simulation is superior to traditional educational methods for skills acquisition. However, their analysis did include the 4 studies with no control group where they only considered the pre and post training scores. The other issue with their findings is that only one study looked at skills acquisition in actual patients, the others either used a checklist type test or checked skills on the simulator the training was done on questioning transference of the skills to the workplace. This type of assessment is reflected in the other studies where only one further study actually looked at changes to performance post training on patients.

The review to this point has focussed on simulations, learning and transfer of learning and although there does appear to be plenty of evidence to suggest learning does occur on a simulator and that it is favourable to established methods, there are questions over the quality of the evidence and perhaps more importantly that of transference of skills learning to the real world. The review now changes tack to consider the issue of competency and assessment which is the main focus of the study being undertaken.

3.8 Competence and competency based assessment

Over 40 years ago Miller proposed a pyramid that identified a framework comprising four stages for assessing clinical competence (see Figure 3.2). The stages range from fundamental acquisition of knowledge to ability to perform a task in practice. The model suggests that the bottom two stages only test cognition, whereas the top two stages concern performance and showing how (demonstration) assess behaviour. The model was later revisited by Glavin & Maran (2002), and Michelson & Manning (2008), (see Figure 3.2).

Figure 3.2 Conceptual Clinical competence pyramids.



Taken from Glavin & Maran (2002: 329), and Michelson & Manning (2008: 610)

In separating out performance and demonstration the model introduces a distinction between the two, a difference Carr (2004) picks up on stating that being able to show in a simulated environment does not necessarily mean that one can "do" in actual practice due to the difference in situations. The implication one might take from this assertion is that the more realistic the simulation is, the closer the assessment outcome is to performance; an idea picked up by Schuwirth & van der Vleuten (2003) who state that one way to increase a simulation's authenticity is to base it on a simulation of reality. Eraut (1994), however, argues that context is of more importance; the greater the gap between the context of use and work context the weaker the inference about capability and performance. So it might be more important to make sure that the context of the simulation is realistic rather than the simulator itself. Picking up on the difference between shows how and does, in the first pyramid above Kak, Burkhalter & Cooper (2001) argue that as medical and paramedical professions are competency based, methods that "show how" such as simulation may be a better method of assessment than performance. The rationale behind this statement being that performance is affected by many causes, not just competence and as an assessment tool performance lacks the reproducibility offered by simulation.

There exists a number of different approaches to describing competence, with most definitions found along a continuum with demonstration of a required behaviour at one end and individual focus at the other, (Meyer in Potgieter & Van Der Merwe, 2002). Whilst there is no one agreed upon definition of exactly what is meant by competence, Vandewater (2004) states that regulatory bodies, such as the COR and HCPC tend to define competency based on the “scientific method” of direct quantifiable outcomes that have certain disciplinary-based responses, so conforming strongly to one side of the continuum indicated by Meyer, (*ibid*) above. One potential issue with an outcomes model is that the emphasis is placed on the outcome; Raffe (1992) reporting concern over this issue that some teachers have as the process itself can be neglected and the sequence and approach to the task often distinguishes an expert performer.

Porter-O’Grady (2001) states that the primary roles of the educational facility now are to help learners select the best form(s) of access to learning to allow the student to meet the competency and then to be able to evidence the competency has been met by the learner.

According to Wolf (2001) there are three separate components of competency-based assessment, which are all equally important;

- There is an emphasis on outcomes; specifically, multiple outcomes, each of which is distinctive and considered separately.
- The belief that the outcomes should be specified to the point where they are clear and "transparent" to all parties.
- The dissconnection of assessment from particular institutions or learning programmes.

There is the possibility that considering each outcome separately could lead to atomising the collection of evidence, (Debling 1992), and the loss of the collective whole that is the competency. To avoid these limitations Gonczi (1993) recommends the use of *integrated* methods (methods which seek to assess some combination of the required knowledge, understanding, problem solving, technical skills, attitudes and ethics in an integrated performance).

3.9 Assessment with Simulation and Virtual Reality

Although the use of simulation is widespread, much of the background literature on simulation is based in the medical or para-medical fields which have embraced this form of learning because of their unique position bridging both academic and clinical environments. As the study being conducted was also in this field the material sourced was almost exclusively medical and para-medically based.

The use of simulation including virtual reality has gained acceptance as a training tool and its use is becoming more widespread. According to Pantelidis (2010) the advantages offered by VR are similar to an interactive simulation in that they allow empowerment of students, encouraging active participation and it is highly motivating for the students. With its acceptance as a teaching or training aid it is natural to consider and adapt its use for assessment purposes, (Ryan *et al.*, 2010). In health practice simulations are now regularly used for assessment purposes, particularly the use of standardised patients for assessments such as communications skills, in Objective Structured Clinical Examinations (OSCEs) (Ryan: *ibid*), and mannequins such as "Resusci Anne" to assess Cardiopulmonary resuscitation **(**CPR) cognitive and psychomotor skills competencies.

Again as with the use of simulation as a tool for learning what is missing in most of the literature is the comparison of simulator performance to real world performance. Authors tend to rely on the premise that a measure of performance on the simulator directly corresponds to a level of performance in real life, but there is little evidence that the results are transferable between contexts. This issue has been acknowledged by Perkins, (2007) who defends this methodology by stating that the impact of simulation training is difficult to measure in clinical practice due to the relative infrequency and unpredictability nature of the timing of what simulators are simulating such as cardiac arrests. Also, that there will be multiple confounding factors when and if they do occur, and therefore a simulator would be a better competency measure as it can be standardised.

The other factor that must be remembered is that most health care assessment is looking to measure competency (shows that they can do), not performance (what they actually do). The issue therefore is more akin to the question how closely does demonstration of competency on a simulator mirror demonstration of competency in real life. If the simulation fails to include all the essential components that exist in real life any lack of fidelity in the system can affect a difference in how they behave on the simulation to the real world as the simulation is no longer a valid representation of the real life scenario. Because of this Moroney & Lilienthal (2008) argue that we cannot accept the fact that any performance in a simulator reflects how an individual may react in a real world environment, as with any tool that can be used for assessment, the VR simulation must demonstrate acceptable levels of reliability and validity. This stance is supported by Thijssen & Schijven (2010) who state that if VR simulators are to be used for the assessment of competence the metrics must be reliable and valid. According to both Winter (2000) and Green & Thorogood (2014) the concept of both reliability and validity are tools of a positivist epistemology. Reliability is concerned with the consistency of the result if the research is repeated, and at a base level validity can be thought as being concerned with demonstrating that what the researcher believes is being measured is actually being measured.

In order to consider these aspects of VR use a literature search was conducted using the electronic databases of Medline and Science Direct in January 2014. The Google Scholar search engine was also used to avoid missing studies that remain unpublished. The Subject Heading terms used were VR, Virtual reality, Assessment, Competency, Reliability and Validity. Articles found that dealt with the assessment of rehabilitation or for treatments such as phobias were excluded.

A number of studies were found considering either the validity and or reliability of VR simulation assessments. Each VR system is unique and hence the reliability and validity only applies to that specific VR system. The review will give a useful background to the evidence base for use of VR as assessment tools in the medical field.

3.9.1 Reliability of VR simulations

Reliability has most often been assessed by test retest, for example Aggarwal *et al*., (2007) tested 19 surgeons and compared the results of the first and second procedure of each subject using a Cronbach's alpha test to ascertain reliability. The authors reported alpha scores of 0·502, 0·623, 0·229, and 0·522 on various components of the assessment and based on these figures the authors reported that the skills assessment being employed was a reliable assessment. Although there are no absolute acceptable levels for reliability Henry *et al*., (1999) report that some authors have suggested that coefficients of 0·60 or higher are acceptable for test-retest reliability although other authors (Kline, 1993) quote 0·80 as the minimum acceptable figure. Pedhazur & Schmelkin (1991) refer to Nunally

(1967) who quote that alphas of 0·50 or higher have reached acceptable levels of reliability, whilst acknowledging other authors suggest 0·70. They suggest that all these figures are acceptable and that the purpose of the research being tested is important, low reliability measures being acceptable in early stages of research, but higher reliabilities are needed if the measure is being used to look for differences among groups. Patel *et al*., (2006) also looked at reliability of a VR simulation using a test-retest method. Again the study reported a low sample size and so has low power, but the study did report good test-retest reliability (r = 0·9, p = 0·0001). Although correlations such as Pearson's product-moment correlation coefficient (r) are regularly used by researchers to test reliability their use should be considered suspect as this method is unaffected by systematic bias, (Hunt, 1986) and it only measures the strength of relationship between two variables; not the agreement, (Bland & Altman, 1986).

Kline (*ibid.*) also specifies a minimum time period of 3 months between assessments and a sample size of at least 100 for this type of analysis. Aggarwal *et al*., (2007) however did not specify the time period between assessments and only utilised a sample of 19 and so was very low powered and therefore the probability of a type II error, failing to reject the null hypothesis when it is false, is high.

Hogle, Briggs & Fowler undertook a study in 2007 to document the learning curve for two tasks using the LapSim surgical simulator, a VR laparoscopic training system. As part of the study they investigated test–retest reliability, on repeated measurements undertaken by students once the learning curve had plateaued. Test–retest reliability was calculated using the intraclass correlation coefficient (ICC) and kappa for each outcome variable on a convenience sampled group of 26 subjects. Results were reported in two different skills, the first looking at camera navigation which reported ICC values of 0·35, 0·52 and 0·72 for the three components. Kappa values for the same three components were 0·4, 0·51 and 0·51. The second skill, lifting and grasping reported the ICC and kappa values of 7 individual components, the ICC ranging from 0·48 to 0·78 and kappa from 0·16 to 0·82.

As with Aggarwal *et al.*'s study this study had low subject numbers, however the author reported that given an alpha of 0·05 and a power of 0·90, a sample size of 25 was needed to estimate test retest reliability, although this statement lacked any cited support. The authors also reported “high” ICCs although Cicchetti, Shoinralter & Tyrer (1985) indicate that when looking at the clinical significance of ICC the acceptable convention is that values of above 0·75 are considered excellent, 0·60 to 0·74 good, 0·40 to 0·59 fair and below 0·40 poor and so many of the ICCs would be classed as fair at best. Kappa values of 0·81 and above are considered almost perfect agreement (Viera & Garrett 2005) and the study’s authors did acknowledge that with all but one measurement having a Kappa value below 0·8 that the protocols being studied would probably not be accepted as part of a surgical training curriculum.

3.9.2 Validity of simulations including VR simulation.

As stated in the introduction, before any assessment tool is accepted for general use it must first be proved to be a valid tool. Historically, validity was defined as the degree to which a test or measuring instrument measured what it purported to measure, but that more recently it has been used not to comment on the instrument itself, but rather the interpretation and inferences derived from the scores of the instrument, (Oluwatayo, 2012). It is this last interpretation that Kaplan, Bush & Berry, (1976) states is important, but they do caution that validity is not absolute, but rather is relative to the domain about which the statement is made. Traditionally it is accepted that there are three fundamental types of validity: criterion, content and construct. These three types of validity were proposed in the early 1970s by the American joint committee of psychologists and educationalists (Kaplan Bush & Berry, 1976) to help minimise confusion when dealing with the issue of validity. These broad categories subsume almost all the forms of validity and are therefore a useful way to organize and discuss the issues of validity.

3.9.2.1 Face Validity

Face validity has been investigated by a number of studies, although whether face validity is an actual empirical measure of validity is questionable, (Downing & Haladyna, 2004), though they do acknowledge that it may be an important characteristic of the tool. Face validity is usually assessed by asking experts in the field their opinion on the tool in question, (Polit & Tatano Beck, 2012). Face validity was assessed by a number of different methods, verbally, rating and questionnaire and in all instances authors claimed that face validity existed as they opinionated that the VR test appeared to measure what it was supposed to measure. Most authors looked at face validity as a multifaceted concept, looking both at the metrics, appropriateness and the reality of the simulator. Without exception all the articles found indicated that the simulation met the criterion of face validity. Face validity was typically assessed by experts on completion of a simulation task. Some such as Bright *et al*., (2012) obtained verbal comments from experts whereas more commonly studies utilised questionnaires, for example Schreuder *et al*., (2009) used a questionnaire that consisted of 27 statements about the simulator asking questions about the realism, training and capacities of the simulator. More commonly, however the number of questions forming the questionnaire to ascertain face validity was much smaller, Brewin *et al*., (2014) and Bongers *et al*., (2014) using 3 and 8 questions respectively, which is more typical of the studies in general. Xiao *et al*., (2014) purported to have measured face validity but used the novice group in the assessment rather than experts and the questions asked were not very appropriate, the three questions being about attractiveness, encouragement to practice and like to own.

While many of the studies laid claim to obtaining face validity about the simulator Bittner *et al*., (2010) perhaps more realistically did not state that it was the simulator that had face validity, but rather that the simulator demonstrated face validity for 2 simulated ERCP cases, so implying the simulations were valid, not the simulator.

3.9.2.2 Content Validity

Content validity was also assessed by a number of studies; this was usually done by questionnaire. Content validity differs from face validity in that it tries to assess if the user believes that the actions and behaviour(s) demonstrated in the procedure are representative of the actions and behaviour(s) of the job in question. One confounding issue with the reporting of content validity from the studies was that often the subjects consisted of both experts and novice users and it was not explicitly if there was variation in opinions between these groups when content validity was considered. Gavazzi *et al's*., (2011) study is typical of this presenting data from experts and novices alike, then stating that both face and content validity had been met as a larger proportion of users (87%) considered the simulator generally useful for training and 90% agreed that the simulator was useful for learning hand–eye co-ordination and suturing. Most studies (Appendix XI) found suggest that VR simulators and simulations looked and felt realistic. A number of studies Bittner *et al*., (2010); Gavazzi *et al*., (2011); Schreuder *et al*., (2009) did question novice users about the system as well as expert users to derive content validity. This is an important consideration in the use of VR simulation as these are the people the VR systems are being aimed at and will be being trained and potentially assessed on them.

Both face and content validity can only be evaluated subjectively. Face validity being the extent to which a test seems to measure what it is intended to measure, and content validity the degree to which a test includes all the items necessary to denote the concept being measured, (Roach, 2006). These types of validity are considered important if you are supporting the take up and use of a particular assessment as is the case for simulation as the simulator needs to appear to measure what the trainers intend it to measure.

Neither face nor content validity can be examined experimentally, and both are considered lower levels of validity, whereas criterion and construct validity can both be objectively examined. The majority of the studies reporting to have investigated validity have focused on the demonstration of construct validity. Construct validity is a quantitative measure of validity testing the ability of the tool to measure the underlying concept of interest to the clinician or researcher. There are a number of different strategies available to examine the concept of construct validity, but researchers using simulators have virtually exclusively utilised the known-groups technique. This method involves administering the measurement instrument to groups expected to differ in their knowledge or performance, e.g. experts and non-experts where construct validity would be assumed if the experts knew more or performed better than the novices on the purported measure, (Wallston, 2005).

Of the studies identified, most compared either two groups (experts and novices) or three groups (experts, intermediate and novices), whereas others compared three groups including intermediary users and one study McDougall *et al*., (2006) used four groups. How groups were divided into expert and novice groups also varied, some using doctors with different qualification periods or experience in the field being studied. Most commonly novice groups were derived from students studying in the field the simulation was covering.

3.9.2.3 *Construct Validity*

Construct validity seeks to identify if the test or system measures the ability or trait it was designed to measure. One example is the ability to differentiate between people who should demonstrate the construct and those that should not. In the case of simulation this would mean comparing groups who are familiar with performing the procedure/task being simulated, the experts, with a group consisting of people who are unfamiliar with the procedure/task and should therefore not be able to demonstrate the construct. This was the method used by all the studies looking at construct validity although the studies do vary in the quality of their assertions, for example, Bajka *et al*., (2009: 1) states that, "Construct validity for HystSim has been established for different modules of VR metrics on a new MMSS developed for diagnostic hysteroscopy." This assertion was postulated despite the overall score not being significantly different for either exercise undertaken, whereas in for the other studies reported in Appendix XI there were clear differences between the groups. In Bajka (*ibid*.) for two of the assessed modules making up the overall score, the experienced group scored significantly higher, but for another module the novices scored significantly higher in both exercises. When the analysis was rerun by the authors again without the safety module where the novices scored higher; this resulted in a significantly higher score for the experienced group than the novice group which led to the statement above; although the authors attempted to justify its removal from the overall metric. Closer examination of the data reveals that there are further issues with the findings from many of the studies in that not all parameters show the expected relationship, for example Bright et al., (2012) found a significant difference between novices and experts on only 2 of the 3 measures. The study by Schreuder et al., (2009), was even more optimistic in its reporting of the establishment of construct validity for the LapSim when only 20 of the 82 parameters measured showed a significant difference between the performances of the 3 groups.

Although most studies concluded from their analysis that their simulation had construct validity, this might not be the case. Brown, (2000) states that construct validity should be demonstrated by both the accumulation of evidence and the utilisation of different strategies to prove validity. This view is supported by Kimberlin & Winterstein, (2008), who state that evidence for validity is built over time with studies looking at a variety of populations. The evidence found on medical and paramedical VR simulations cannot be said to be convincing for individual simulators, except perhaps for the LapSim which has a number of validity studies looking at it (Appendix XI). Most simulators only had one study investigating face and/or content validity undertaken on them, so the wealth of evidence for each simulator is lacking. However, the evidence for the validation for virtual simulation as a whole might be more convincing given the positive nature of the studies as a whole.

3.9.2.4 Criterion Validity

Criterion validity is considered as being one of the most straightforward types of validity. It is usually tested by comparing the results of the outcome measure or target test to the results from an established method, *i.e.* how well scores from one measure predict scores on another measure of interest (the criterion), (Salkind, 2010). This type of validity can be examined by the subjects having both tests at the same time (concurrent validity), or by giving the tests one after each other to determine whether one predicts the findings of the second test which would be administered at a later time (predictive validity), (Roach, 2006).

The primary problem with this form of validity is that it requires an established gold standard which in this case would be how subjects perform in real life, and as stated earlier there are very few situations in health care where a standard against which you can measure reliably exists. This reason probably explains why only one study, Prystowsky *et al*., (1999) has reported this measure for VR simulation. Prystowsky *et al*., (1999) compared VR intravenous insertion with practice intravenous insertions on fellow students, but found no correlation in performance between the two and so stated that no concurrent validity existed.

3.9.2.5 Other thoughts on Validity

One study (Tsai *et al*., 2008) as well as looking at validity and reliability of simulators also considered stability of measurement over time. Subjects were asked to repeatedly carry out a procedure and it was noted that after a few practice trials which took a long time and gave varied results that the time and error frequency soon decreased and became stable. Although only substantiated by one study this finding is not unexpected, but it does highlight the importance of familiarisation with the system in order to produce accurate results.

Another issue raised by Stunt *et al*., (2014) is the sheer number and variety of simulations currently in use. Although they only considered surgical skills training simulators they identified four hundred and thirty three commercially available simulators of which only 35 (8%) been tested for validity. This, potentially is an issue for the educational users and trainers using the simulators, however, the problem may be even more pronounced than that. Many simulators can be used to undertake multiple simulations, for example VERT can be used to simulate electron setups, different types of x-ray setups and quality assurance checks. Each simulation is different to the other and uses different degrees and types of interaction and to undertake a validity study on one of these aspects does not necessarily confer validity on the other types of simulation. Research Design & Methodology

Chapter 4 - Research Design and Methodology

4.0 Chapter Summary

This chapter provides a detailed explanation of the research design and describes the methods utilised to enable collection and analysis of the data. An overview of the mixed methods approach is given first followed by details of the pilot and main study. Both consisted of a quantitative element, a crossover observational study looking at electron setups performed on VERT and on a LINAC. This chapter also includes details of questionnaires that were used in the procedure. This was followed by a qualitative phase which utilised focus groups to generate data. The development of the method between the pilot study and main study is explained as are the procedures utilised in the analysis of data, as well as the steps that were taken in order to ensure the validity of the study.

4.1 Introduction and epistemological stance

In order to capitalise on the strengths of both quantitative and qualitative methods the study was designed to combine both methods and so the study can be described as a mixed methods design. According to Bryman (2007) the term ‘mixed methods’ can be defined as a procedure of collecting, analysing, and integrating both quantitative and qualitative data within a single investigation. The premise behind the use of mixed methods was that in using both quantitative and qualitative methods the data provided would provide a more complete answer to the research question than would have been possible through collection using a single method, testing validity of the method and strengthening confidence in the results, (Smeby, 2012). Brannen & Halcomb, (2009) acknowledge this asset of mixing methods stating that the strength of the mixed method design is that if the data collection methods are well combined the limitations of one method are compensated by the strengths of the other. Howe (2012) builds on this concept further stating that mixing both qualitative and quantitative methods together allows researchers to triangulate on causation.

Despite this rationale for the use of a mixed methods design a number of potential issues arise from using this form of approach, which can be referred to as the 3 P’s, Paradigms and philosophical assumptions; Pragmatism; and Politics, (Brannen, 2005).

The first issue raised is that of the research paradigm as it combines elements of the positivist paradigm of quantitative research with the constructivist paradigm (or variants thereof) found in most qualitative research. Hall (2012) contends that mixed methods only fit within transformative and pragmatist worldviews. Some researchers (Johnson and Onwuegbuzie, 2004; Johnson, Onwuegbuzie & Turner, 2007) allege that a third separate research paradigm exists which includes mixed methods research, based philosophically in pragmatism (the second P), and contend that quantitative and qualitative designs can work together. This would form a trinity of otherwise distinct approaches, and fundamental to this philosophical approach is the belief that the research design should be driven by the research question(s) not the method or paradigm, (Muncey, 2009). This potentially is another issue as Brannen *op cit.* states that, although researchers initially justify their approach in terms of tailoring methods to a particular research question(s), in practice the practicalities of the research process diverts the researchers' original intention to focus on justification by outcome. Finally, under the pragmatic heading, there is the issue of feasibility as the researcher must be conversant with both quantitative and qualitative research methods in order to undertake the research and most are not, tending to specialise in one type of method of enquiry.

The final challenge to mixed methods research is political, with it being alleged (Cameron, 2011) that certain disciplines have set traditions. These traditions favour both the production and publication of particular types of research which may lead to researchers tending to report only quantitative or qualitative aspects of their work, especially in relation to word length limitations present in many academic publications.

A number of different types of mixed method design exist, the method utilised by this study is best characterised as an exploratory sequential design (Creswell & Plano Clark, 2010) as it consists of two distinct stages, a quantitative stage of data collection preceding a qualitative stage of data collection. The purpose of the qualitative data was to attempt to explain the initial quantitative results. The emphasis on the quantitative aspect has led to this type of study often being referred to as QUAN → qual.

With the emphasis being on the quantitative aspect of the research ontologically this is a study that accepts that there is a scientific single object reality, although knowledge around this might be socially constructed. As the objective is a single reality it should be independent to the researcher's perspective or belief and as such the study has taken a controlled and structured approach in constructing appropriate hypotheses and by adopting a suitable direct investigation.

This argument falls into the positivist paradigm, which has a commitment to an empirical epistemology, where researchers try to remain detached from the participants of the research by creating a distance, which is important in remaining emotionally detached, in order to base conclusions on reason rather than feeling and personal experience, (Cruickshank, 2012). Because of this detachment it is important in positivist research to work from hypotheses which are tested by statistical and mathematical techniques in order to uncover a single and objective reality.

The advantages offered by this design are that it is relatively straightforward to conduct: the two stages are discreet and so the data is collected one type at a time, (Creswell & Plano Clark, 2010). The authors (*op cit.*) also state that this type of mixed method design appeals more to researchers with a quantitative background, as in this case, because of its marked quantitative orientation.

4.2 Study Phases

This study was comprised of three major phases. The first consisted of a pilot study and material development, while the second, the pre-experimental phase consisted of pre-experiment subject selection and randomization process, the third and final stage consisted of data collection for the main study.

4.3 Phase 1: The Pilot Study

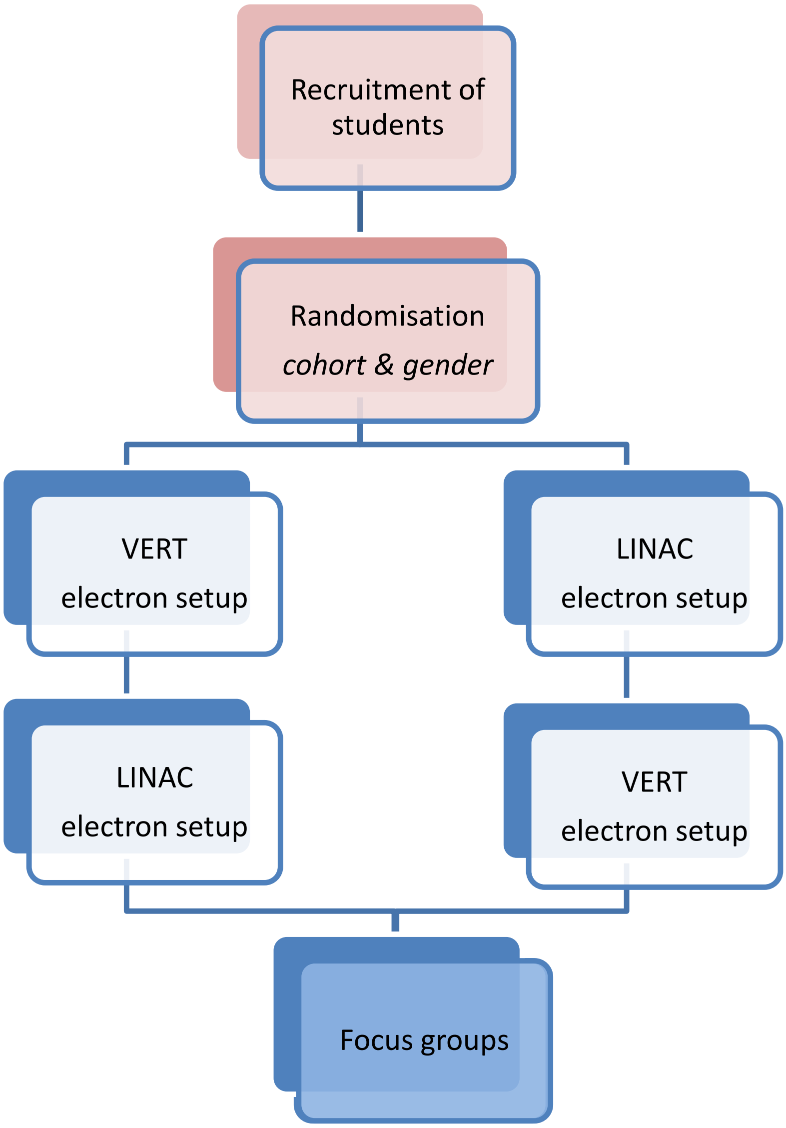
The pilot study was conducted in 2010 with the participation of radiotherapy students from all three years of the radiotherapy course at City University, London; the design of which is shown in *Figure 3.1.* Ethical approval for the pilot study was granted by City University London, School of Health Sciences Research Committee. Approval letters can be found in Appendix I. Students were randomised using a stratified randomisation procedure based on gender and cohort into one of two groups. Stratified randomisation is often undertaken in order to ensure both a balance in sample size and covariates across the two groups, (Suresh, 2011). This aspect was considered important given the relatively small sample size of the study and the small number of males present within the sample.

Once allocated to a group students were asked to complete an electron set-up on both the VERT system and a linear accelerator, the order being decided by the random allocation to one of the two groups. Both setups were undertaken using a Varian unit, however, the units were different as the clinical unit available at the time was a Clinac 6 which is not modelled in the VERT package. A small number of students were then invited to take part in two focus groups.

4.3.1 The electron set-up

The VERT system comes supplied with a patient file containing ten electron treatment fields marked on the patient. All the electron fields provided utilise a 10cm square applicator and a 10cm by 7cm insert, (see section 2.2.1 for details about electron setups). A number of fields of varying difficulty were transferred onto an anthropomorphic phantom using an electron end plate to approximate the measured standoff obtained with VERT. Four experienced radiographers who were working in higher education and had previously used VERT then independently completed all the setups utilising VERT version 2.6, and a Varian CX linear accelerator using the anthropomorphic phantom. The electron setup with the least variation of stand-off between the VERT and LINAC setups was then selected for comparison in the pilot study.

Figure 4.1 Flow chart of the Pilot Study



4.3.2 Electron measurements

Similarity of setup on VERT and the linear accelerator was established by looking at the amount of standoff in each corner of the field and the treatment distance (FSD). In VERT this was straightforward as, at the end of each student setup, the measurement tool was activated that provides the standoff distance in each corner. The treatment distance was measured using the FSD distance light. For the LINAC measurements a ruler was inserted in each corner of the field and the distance to the skin measured. The treatment distance (FSD) was also recorded using the distance light. When undertaking the pre-experimental stage of the study this method was again repeated to measure the respective distances.

4.3.4 Focus groups

Two focus groups were held at City University, the aim of which was to compare the methodology of the two setup methods. Students were asked about their experiences on both systems and in order to ascertain the strengths and weaknesses of the methodologies employed. Focus groups were used as they allow social interaction amongst the participants which can stimulate the production of ideas through dialogue. This allows rich data to be produced as participants operate within a social context, reacting to the ideas of others, and influencing others as they operate within a social context, (Walden, 2012). This therefore allowed substantial data to be obtained relatively quickly in comparison to other methods, such as interviews. Another benefit of focus groups is their ability to be used with other methods making them well suited to mixed methods research, (Walden *op cit*.; Hennink 2014).

Additionally, focus groups have been shown to be an effective strategy when used in consumer research, (Threlfall, 1999). Although there is debate over whether students are consumers or not, certainly there is more of a move to consider them as such in Higher Education. The legal relationship between Universities and their students is changing, a recent report stated that institutions could be failing to meet their legal obligations to students under consumer protection legislation, (OFT, 2014), so implying that legally they are considered as consumers. Because of this University students today are more empowered than ever, with many quality metrics used by HEI's being student centred measures. This leads Palfreyman (2013: 109) to describe students as "empowered but not necessarily well-informed consumers." And, as consumers, students have a right to opine about the introduction of technology and learning methods such as that proposed in this study in their programme.

4.4 Phase 2: The pre-experimental stage

Following the pilot study the results were presented at the 4th International VERT User Meeting. Feedback obtained from the pilot study and from the delegates at the meeting helped inform changes incorporated into the main study.

4.4.1 The electron set-up

During the pilot study it was established that both the FSD and distance to the patient in each corner of the beam on both VERT and the LINAC approximated each other although differences did exist. Differences might have existed due to student performance, but it was also recognised that although the setups on VERT and the LINAC were similar there were differences due to the patient outlines being different. For the main study it was decided to introduce identical set-ups on identical treatment units in both arms of the trial so as to ensure that any differences that occurred between the two set-ups arose solely as a result of the student's action.

An Alderson RS-111 anthropomorphic phantom was sourced that could be used in the linear accelerator arm of the trial. The phantom extended from the neck to just below the diaphragm and consisted of moulded tissue equivalent material representing a male skeleton 5 foot 9 inches tall and weighing 162lb. The phantom was then scanned using a Siemens SOMATOM Emotion dual slice computerised tomography (CT) scanner using a slice thickness of 3mm with a slice interval of 3mm. The data was then exported using a DICOM file format (Standard file format for the transfer and storage of information in medical imaging) to a Varian Medical System's Eclipse radiotherapy treatment planning workstation where the slices were outlined. This was done by firstly using the automated outlining tool and then manually adjusting the outline to ensure an accurate outline to the phantom's contour. This method of outline production was in essence identical to that used in radiotherapy on real patients which has proven clinical accuracy.

Once complete the file was once again exported as a DICOM file and sent to Vertual (the manufacturers of VERT) to have electron fields placed on the outline. The DICOM file was returned with electron fields positioned on it and two were selected for use in the study. The two electron fields were then placed onto the anthropomorphic phantom using measurements from known anatomical landmarks. The position of the field was independently checked for accuracy by a second radiographer.

4.5 Phase 3: The experimental stage (QUAN)

4.5.1 The Setup

Once identical "patients" had been created for both VERT and the LINAC further standardisation occurred. The VERT simulation was positioned at a "usual" start point, the couch close to the floor and out from the machine to simulate a position that would make it easy and safe to get the patient onto the couch. The patient data was uploaded and the electron field placed on the patient and the electron applicator and cut out loaded into the simulation.

This was then loaded into the virtual presenter that allowed the setup to be set identically for each subject when they came into VERT. The machine parameters were then recorded which allowed the researcher to produce an identical start point on the LINAC and also ensure that the patient was positioned as on VERT. Full details of the similarities and differences between the two simulations can be found in Section 2.4.

4.5.2 Stratified Randomisation

Stratified randomisation was undertaken of the sample based on three variables, ITQ, cohort and gender. Stratified sampling has a number of important advantages over simple random sampling. Firstly it ensures that there is an equal balance in sample size across the two groups. Its main feature though is to ensure a similar distribution between treatment groups of important variables that are thought to influence outcome, Kahan & Morris (2012), in this case ITQ, cohort and gender. All these variables have the potential to affect performance and therefore it is important that they be considered at the planning stage so that they don't affect the analysis later as the groups are balanced in design.

The rationale for each variables inclusion in the stratification process as a variable that might affect outcome is briefly discussed below.

As stated in section 3.3.1 realism was an issue in the pilot study for subjects when using VERT. The concepts of immersion and presence are discussed in sections 3.5.2 and 3.5.3 of the literature review. Because this issue was raised by the pilot study's focus groups it was decided to introduce two instruments, the first to measure immersive tendency and the second presence to investigate if this had an influence on student performance.

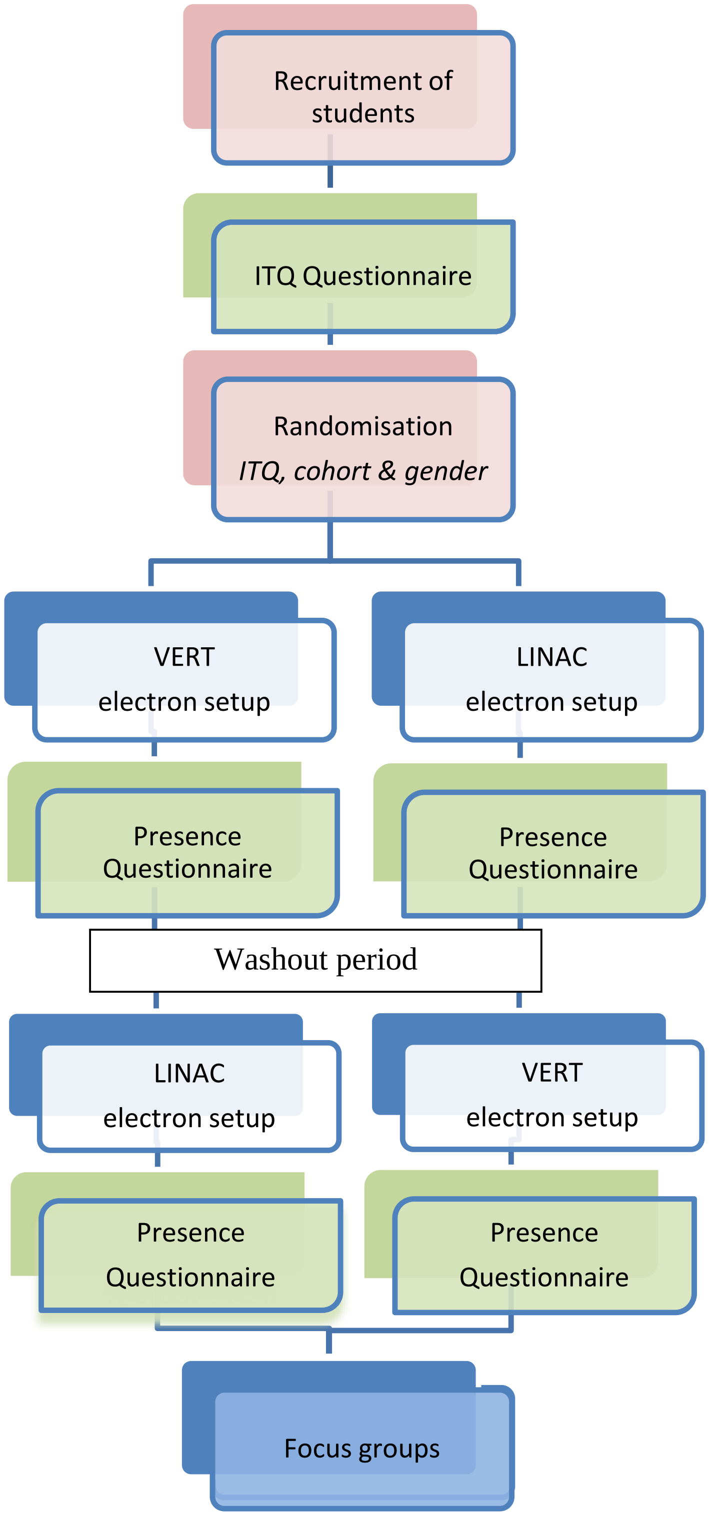
Immersive tendency was measured using the Immersive Tendency Questionnaire developed by Witmer and Singer in 1996 and revised in 2004. The questionnaire (Appendix IV) was sent to all subjects who agreed to take part in the study prior to undertaking the electron setups. The questionnaire was also modified to include some simple demographic questions. Immersive tendency has been shown to have an indirect effect on a student's performance in a VE, mediated by presence (Jerome & Witmer, 2004). As presence could only be measured after the simulation immersive tendency as measured by the questionnaire was used as a surrogate measure with which to randomise subjects.

Spatial ability is a mixture of three different elements each based on a different process required to solve problems, spatial perception, mental rotation, and spatial visualization, Voyer, Voyer & Bryden (1995). A large amount of literature exists that suggests that spatial ability is affected by gender with some studies implying this is because of nurture rather than nature. Whatever the reason men tend to outperform women in most studies, (Mäntylä, 2013; Pietsch & Jansen, 2012).

Students within different cohorts will have been exposed to different levels of clinical experience and as such one would expect students in an advanced cohort to perform at a more advanced level than those present in cohorts behind them. This expected difference is the basis for the measurement of construct validity which was discussed in section 3.9.2 and the results of the various cohorts in respect to construct validity is looked at further in section 7.6.

The final design of the study is shown over in Figure 4.2.

Figure 4.2 Flow chart of the Main Study



4.5.3 Observation study

Students were observed doing identical setups on both the VERT system and a LINAC, this type of study design being classified as a cross-over design. Because each subject undertakes both aspects of the study each subject serves as his/her own control, (Wellek & Blettner, 2012), this removes “patient effects” so reducing variability and increasing the precision of the estimation which means that smaller sample sizes are typically needed for this type of study, (Maclure & Mittleman, 2000). Cross-over studies operate under the assumption of no carryover effect, an effect that carries over from one experimental condition to the other and affects the result of the second experimental condition therefore introducing bias into the study. In order to reduce the possibility of this effect happening a cross-over study requires a washout period between each experimental condition, (Senn, 2002).

In order to standardise data collection, (a potential area of bias identified within the pilot study), one observer was used throughout the whole process using a data sheet designed on the experience gained from performing the pilot study for recording of the variables, (see Appendix V). In the pilot study on completion of the setup by the subject the distance at the centre of the field and at the four corners was recorded, averaged and the standard deviation calculated in order to provide a measurement that could be used to compare the set-ups. For the main study these measurements were again repeated and used to calculate the dose variation at a depth of Y. Additional data collected at the time of the setup were light beam to field mark fit, if there was a machine crash during the set-up and whether the student was moving the machine parameters singly or in combination.

The use of VERT was also altered to more closely represent real life in that students were asked to perform the main study in tracking mode which wasn’t used in the pilot study. Tracking mode relays signals to the system about the orientation of the user’s point of view, letting the user roam around within a physical space. The locality of the person can then be detected along with his direction and speed and the images and sound are altered accordingly.

Another addition to the main study not undertaken in the pilot study was a time-lapse video of the set-ups in order to try and evaluate if there was a difference in the subjects approach to the set-up as well as the end point which was the sole focus of the pilot study.

Because of the issues in videoing the movement of the users in both systems and that of image quality on the VERT system, rather than video the subjects using the equipment, a camera was setup to record the beam parameters either from the control panel in the case of the LINAC and from the beam parameter section of the VERT projection. Time lapse photography was used with images being taken every ten seconds during the student's set-ups. This data was analysed later and the setup parameters noted at the end of each 10 second period. This data also provided the time taken to perform the set-up.

Both the VERT and clinical setup was conducted using a Varian 2100IX linear accelerator, utilising identical Varian handsets to manipulate the unit.

A number of possible instruments exist that could be used to measure presence. Objective measures were not considered for this study, primarily due to the fact that their main use is in simulations that elicit specific physiological responses such as the stress experienced when flying a plane that is experiencing mechanical failure. There were other considerations such as the need for free movement which can be limited by these devices, but was considered essential in this type of simulation, the cost of specialised equipment and limited proven reliability, (Insko, 2003). Task performance measures such as completion time and errors were excluded as these measures were being utilised to establish similarity in performance between the systems and adding secondary tasks would make the assessment unrealistic and might interfere with the primary goal of the study.

A number of subjective measures to measure presence exist, such as hand held sliders, the Turin test and questionnaires. Of these the slider was rejected as this measure requires the subjects to hold the slider during the setup which may interfere with the interaction with the handset as it requires two hands to operate effectively. The Turing tests could have been utilised as they are sensitive, cheap, and easy to use, but have only been used in a limited number of previous presence studies (Van Baren & Ijsselsteijn, 2004) and are prone to experimenter bias, (Wissmath, Weibel & Stricker, 2008). Most previous research had utilised questionnaires, the evidence for the reliability and validity of these instruments far exceeds that for the other types of measurement.

One of the most commonly used presence questionnaires (PQ) was that designed by Witmer and Singer in 1998. This questionnaire had established reliability and sensitivity, (Van Baren & Ijsselsteijn, 2004). The authors (*ibid.*) report that the questionnaire has both content and construct validity although they admit that at the time of publication there was no research that directly compared the PQ scores of a single group of participants across different VEs. Youngblut & Perrin (2002) confirm that the Witmer-Singer PQ is one of the most widely used presence questionnaires and in a comparison with other questionnaires (Table 4.1) identified several important advantages of this tool compared to others. Another advantage of this questionnaire over others identified was that it could be used with no modification in the real world, (Usoh *et al.*, 2000).

Table 4.1 Presence questionnaires.

Questionnaire name	Face validity	Vary with related factors	Stable for unrelated factors	Comparison with other types	Consistency across studies
Hendrix's questionnaire	-	-	-	-	Y
Biocca's questionnaire	Y	-	-	-	-
ITC-SOPI	Y	-	-	-	Y
Sa's questionnaire	-	-	-	-	Y
SUS questionnaire	Y	-	-	Y	-
UCL questionnaire	-	-	-	Y	-
Witmer-Singer PQ	Y	Y	Y	Y	Y
Mania's questionnaire	-	-	-	-	Y
Tromp's questionnaire	Y	-	-	-	-

Taken from Youngblut, (2003): 12.

To ensure accuracy of data used in the analysis all data generated from the questionnaires and videos were entered into an Excel spreadsheet using a double entry method. Two identical blank data entry spreadsheets were created. Data were then entered by the researcher onto the two different spreadsheets on separate occasions and a third spreadsheet was produced to compare and verify the accuracy of data. Any differences highlighted by the third spreadsheet were clarified by comparing the original questionnaire/video with both data entry spread sheets and correcting the relevant data entry spreadsheet.

4.6 The experimental stage (qual)

Following on from the information gathered in the pilot study, one question to answer was if there was another need for a qualitative element to the main study. The pilot study focus groups were helpful as they raised issues that were not apparent in the quantitative element of the study and informed changes to the main study such as the introduction of the presence questionnaires. Significant changes were also made to the experimental section of the study, such as moving from both 2D and 3D used in the pilot study to 3D with tracking. The data provided by the quantitative part of the study could only identify similarities or differences in performance in the simulations. Another aspect, the acceptability of each method to the students and what the users felt about each experience and their insight and attitudes to their use could not be measured this way and so it was felt that qualitative research in the form of focus groups still had a role to play. This stance is supported by Tuli (2010) who states that qualitative methodologies are inductive, that is to say are more oriented toward discovery and deeper understanding of the research problem.

The qualitative method selected was that of focus groups, a type of group interview that Rabiee (2004) states can provide information on ideas and feelings individuals have about issues on a certain topic, as well as being able to illuminate differences in perspective between groups and individuals. Although the number of focus groups required is debated by authors Stewart *et al.,* (2007) observed that rarely are more than 3-4 focus groups conducted when the method is used in the social sciences. Krueger (1994) recommends a total of three to five groups per project for a simple research question, the required number of focus groups may only be three or four, so supporting the number utilised within this study.

4.6.1 Qualitative Procedures

As shown in Figure 4.2 the focus groups were performed after the quantitative part of the study and the subjects had experienced both the VERT and LINAC electron set-ups. A focus group guide (Appendix VI) was produced in advance of the focus groups taking place and used by the facilitator in each focus group. All the focus groups were held in private rooms on the students' hospital sites and run by the same facilitator. The focus groups were recorded using both a digital audio recorder and video recorder, the positions of which were checked for sound quality and coverage prior to commencement. Participants were arranged in a semicircle as suggested by Holloway & Wheeler (2009) in order to encourage discussion with the facilitator in the centre

Each group consisted of a heterogeneous mix of students (age, gender, and cohort) from one clinical site and had a group size of between 6 and 8. Each focus group lasted for approximately 40 minutes

Notes were also taken by the facilitator during the meeting. The audio recordings were then transcribed verbatim by the researcher using the transcription software Express Scribe Pro and typed into Microsoft Word. After typing accuracy was checked using the videos, notes were then added to the transcribed documents based on the facilitator's notes and the participants' actions noted on the videos such as pauses, gestures and facial expressions. After checking the transcribed notes the documents were then anonymised. This process lasted approximately 5 months.

4.6.2 Quantitative Analyses

The quantitative analysis was undertaken primarily using the Statistical Package for the Social Sciences (SPSS) version 21. SPSS is a widely available software package that was used as the main tool for the quantitative analysis due to the researcher's familiarity with the software; the researcher having attended a number of courses on the software's use in analysing data and having used it for analysis in previous publications. Medcalc 14.12.0 was used for the Bland Altman analysis as SPSS currently lacks this feature. Medcalc is a more intuitive programme than SPSS, whose use is aimed at biomedical researchers. The software is owned by the researcher for home use.

Statistical significance was set at $p \leq 0{\cdot}05$ throughout the study, that is to say that it was accepted that there was a 5% chance that any given significant result is false (Type I error). Analysis of data was undertaken using both paired and unpaired tests in order to consider all the subject data. The parametric analysis of interval data relies on two assumptions, firstly that the data is normally distributed and secondly that there is homogeneity of variances. Testing for distributional assumptions was a two-step process. Prior to analysis the Shapiro-Wilk test was used to test if the data significantly varied from the normal distribution. For data that had a normal distribution parametric tests were utilised, if there was significant variation ($p \leq 0{\cdot}05$) from normality non-parametric tests were used. The second assumption (homogeneity of variance) is that the variance in the different groups is the same. Homogeneity of variance was tested using the Levene's test, a powerful and commonly used test for this type of analysis, (Hill & Lewicki, 2005). In the instances where there was observed deviation from the assumption of homogeneity the Welch's t-test was performed which takes unequal variances into account.

Whilst most data were direct measurements from the study, dose variance needed to be calculated. The dose at each corner of the field was calculated using the formula on page 24 of this thesis. The effective SSD was assumed to be 80cm in keeping with figures suggested by Min-Tae *et al*. (2014). The obliquity correction factors (θ) were also taken from a table provided by Khan & Gibbons (2014) using data based on measurements taken at the completion of each setup and an arbitrary depth 2cm was assumed.

4.6.3 Qualitative Analyses

Creswell & Plano Clark (2010) state that for mixed method designs quantizing the qualitative data for analysis is often utilised giving rise to magnitude coding. Therefore, the first step in the analysis of the qualitative data was to produce a concordance list from the focus group data using the software Concordance®. The concordance list identified the word frequency and also allowed limited keyword in context viewing. This allowed the researcher to identify frequently used key words such "pendant" and "glasses", and because of the context report also view the settings the words were being used in, so helping prepare for the next stage of analysis. Another useful feature was that whilst reviewing the list of words it became more apparent before undertaking the detailed analysis when synonyms had been used when talking about the same objects by different subjects/focus groups. For example some subjects used the term "goggles", others "glasses"; but they were taking about the same physical object and putting these two terms together showed the importance of the glasses to the students given the frequency it was mentioned in the focus groups.

Qualitative data were transferred into pdf format for manual coding and identification of themes, the starting point of which was the concordance list. The second step of the analysis was to read each of the transcripts thoroughly then two of the focus groups were re-read again and keywords and phrases were recorded that appeared significant to me in relation to the aims and objectives of the research.

The coding system utilised could best be described as descriptive coding, which according to Saldaňa (2012) is appropriate to virtually all studies and in particular researchers coding data for the first time. According to Elo & Kyngäs (2008) there are two possible approaches to analysis depending on whether the categories are derived deductively or inductively. In the deductive approach the researcher begins with predetermined key words or categories which are based on earlier published work in the area and all coding is done using these to undertake the data reduction, (Kondracki, Wellman & Amundson, 2002). The inductive approach, which was used in this study, is when the researcher avoids using preconceived categories; instead the categories are derived from the data emerging from frequent or dominant categories inherent within the raw data. This type of approach is common in health and social sciences research, (Thomas, 2006).

Then in accordance with the technique suggested by Creswell (2002) and incorporated modifications suggested by Mayring (2000) the categories were reviewed and revised to reduce overlap and redundancy within the categories. At this point the concordance was also helpful as it was used as a crosscheck to ensure I had not missed any keywords or phrases that were frequently used by the interviewees. The process was then restarted using the revised categories on all the focus group transcripts.

4.7 Ethical considerations and Approval

Ethical issues are present in all research and therefore acting in an ethical way is important when considering any research. Orb, Eisenhauer & Wynaden (2001) proposes that ethics is related to doing good and avoiding harm and as such the researcher must ensure that harm is reduced or removed through application of ethical principles in order to protect participants. Johnson (2011: 40) takes the idea further stating that research must be honest accurate, quoting the developments being considered at the 2nd World Conference on Research Integrity that research should be carried out using four basic principles:

1. *Honesty* in all aspects of research,
2. *Accountability* in the conduct of research,
3. *Professional courtesy and fairness* in working with others,
and
4. *Good stewardship* of research on behalf of others.

In accordance with the UEL guidelines to any proposed research involving ethical parameters that relate to the Health Department approval was first sought through the Integrated Research Application System (IRAS) as the University Research Ethics Committee (UREC) is not a recognised authority by the Health Department. The IRAS form was completed and submitted to the Joint Research Management Office for St Bartholomew's Health NHS Trust, Queen Mary University of London and East London NHS Foundation Trust as elements of the study were to take place at St Bartholomew's hospital. After consideration and various communications the Research Office decided that NHS REC review was not required as the work did not include any NHS patients or staff and the focus was on educational development. An approach was then made to the University of East London University Research Ethics Committee.

Ethics approval was granted by the University of East London University Research Ethics Committee for the investigation on the 20th October 2012. Approval was also needed from City University, London Research Ethics Committee as part of the research was to be carried out on students on its premises by a member of staff. Approval for the research was confirmed on the 24th of October 2012. Letters confirming this are present in Appendix I. The Letter of invitation, Subject information sheet and Consent letter for the study can also be found in Appendix III.

Two potential risks were identified with undertaking this research. Firstly the possibility of students experiencing simulator sickness during the VERT set ups, symptoms of which are similar to motion sickness. Secondly the potential damage to hospital equipment through student use, such as collision of equipment with each other or the phantom. The risk assessment form associated with the proposed work and included as part of the UEL ethics application is included in Appendix VII. Both these risks were considered low; previous research on simulator sickness (Flinton & White, 2009) had identified that side effects tended to be minor such as eyestrain or headache which tended to be relieved on removal of the glasses. As regards the risk of equipment damage most modern radiotherapy equipment has sensors built into it to protect key components from coming into contact with objects and the observer undertook the secondary role of watching for the possibility of collision and interrupting the set-up to protect the equipment from possible damage.

Chapter 5 - Results

5.1 Background

This chapter presents the results of the study and considers the analysis of the quantitative data. Analysis of the qualitative data is undertaken in the next chapter.

A total of 57 participants (43 female, 14 male) were recruited from student radiographers undertaking their degree at City University during the 2012/13 academic year, which represented approximately 69% of the eligible student population at the time of the study. One student after completing the initial questionnaire did not attend their scheduled appointments and ignored repeated attendance requests for subsequent days and a second went on maternity leave and so both were excluded from the study. Furthermore all the students barring three performed the VERT part of the study in 3D. Given the small number who performed in 2D these data were also excluded from the analysis giving a final sample size of 52.

Due to a number of issues the data set is not complete, on two occasions the VERT hand-pendant interface had connection issues, the video recorder also malfunctioned on two occasions affecting a number of students undertaking the LINAC setups and a number of students failed to return their presence questionnaire despite follow up. Because of these issues and criticism of the crossover design when there is missing data (Simpson, Hamer & Lensing, 1999), analysis of the quantitative data, where possible was undertaken twice. First the data from the first assessment undertaken by the subjects will be analysed using unpaired tests effectively treating the data as though the study was a parallel design. The advantage of this is that the data is easy to analyse and interpret, however the comparison undertaken is between subjects and therefore will lack power compared to the crossover analysis. The second analysis will consider paired data from subjects who completed both the VERT and LINAC setups. This analysis will inherently have more power, but is more difficult to analyse due to the possible unequal carry-over effects and the issue of missing data mentioned above. By undertaking both types of analysis all the useful data will be utilised, so meeting what Newman (2014) describes as a fundamental principle when dealing with data which has missingness.

The main issue of having missing data is that bias becomes an issue if the missing data is systematically different from those that are present, so introducing a bias whose magnitude is difficult to predict, (Matthews & Henderson, 2013). Within this study missing data existed for a number of reasons. Only one student dropped out of the study once it had started and the main reasons for data exclusion were due to failure of the video recorder and technical malfunction with the VERT suite.

According to Howell, (2008) missing data can be classed according to a taxonomy developed by Rubin which classifies missing data as being either Missing Completely at Random (MCAR), Missing at Random (MAR) or Missing Not at Random (MNAR). According to Howell (*ibid*.) equipment malfunction, which accounts for the vast majority of missing data within this study can be classed as MCAR. This type of missing data does not introduce bias into the study as the probability that the variable value is missing does not depend on the observed data or on the missing data values. In this case the only effect the loss of data has on the study is the loss of power within the design and associated larger standard errors, but the estimated parameters are not biased by the absence of data. Due to the missing data being considered as MCAR the subsequent analysis was not adjusted for the missing data. Because of the deletion of data the Paired study effectively became a Complete Case Analysis whereas the unpaired data analysis was an Available Case Analysis.

In this section each of the separate objectives of the study listed below will be looked at.

- To discover if Immersive tendency predicts the feeling of presence in both simulators.
- To assess if Presence will be higher on the LINAC simulation compared to the VERT simulation.
- To discover if Immersive tendency predicts the feeling of presence in both simulators.
- To assess if Presence will be higher on the LINAC simulation compared to the VERT simulation.
- What student characteristics moderate the presence scores?
- Do the outcomes of the two simulated setups utilising different equipment agree with each other?

- What factors moderate the simulation scores?
- Do participants utilise the same cognitive process on both systems?
- What do students think about the appropriateness of both methods as an assessment tool?
- What do students think about the use of VERT as a method of learning?

5.2 Demographics

5.2.1 Unpaired data

Of the 52 subjects 22 undertook two electron setups on the LINAC first and 30 subjects the two identical electron setups on VERT first. A breakdown of the demographic data of the two groups can be seen in Table 5.1.

Table 5.1 Group descriptives.

	N	Age Mean and S.D.	Male to Female ratio	Cohort Ratios Years 1, 2, 3
VERT	30	25·0 (6·2)	9:21 (1:2·3)	5:17:8 (1:3·4:1·6)
LINAC	22	23·3 (6·8)	5:17 (1:3·4)	3:11:8 (1:3·7:2·7)

Prior to undertaking either simulation students were asked to complete the Immersive Tendencies questionnaire (ITQ). Table 5.2 below shows the ITQ results of the two groups, the purpose of the ITQ questionnaire was to establish individual differences in the ability to experience presence prior to randomisation to one of the two arms of the trial. Despite the drop out of subjects during the study the figures in Tables 5.1 and 5.2 indicate that the two groups in each arm of the trial have similar ages and ITQ scores. Statistical tests revealed that this small difference was not significant. (Age $t = -0{\cdot}934$, $p = 0{\cdot}3527$), ITQ total $t = -0{\cdot}411$, $p = 0{\cdot}6827$, subscales ($t = -0{\cdot}476$, $p = 0{\cdot}6363$; $t = -0{\cdot}189$, $p = 0{\cdot}8510$, $t = -0{\cdot}688$, $p = 0{\cdot}4949$).

Table 5.2 ITQ Results.

	N	Total Mean & S.D.	Focus Mean & S.D.	Involvement Mean & S.D.	Games Mean & S.D.
VERT	30	74·0 (15·2)	31·2 (5·4)	28·1 (8·5)	5·3 (3·1)
LINAC	22	75·8 (16·3)	32·0 (6·2)	28·5 (8·3)	6·0 (3·9)
Overall	52	74.7 (15.5)	31.5 (5.7)	28.2 (8.3)	5.6 (3.4)
Normative data	152	76·7 (13·6)	40·3 (6·1)	26·5 (7·2)	6·2 (3·2)
Test of difference		t = 0·8827, p = 0·3785	t = 9·1270, p < 0·0001	t = 1·4122, p = 0·1594	t = 1·1486, p = 0·2521

Normative data taken from Witmer & Singer (1998)

The mean participants' score for this study was within the normal range published by Witmer & Singer (1998) for the immersive tendencies questionnaire for all scores except Focus which produced a significantly lower score compared to the normative data. This indicates that subjects in this study had a lower state of mental alertness than the normative group which would affect their ability to concentrate on enjoyable activities and block out distractions. Subjects scored higher in this study meaning they had a greater tendency to get involved in passive activities such as reading or watching TV whereas the games score was slightly lower indicating less frequent video game playing and involvement in video games. Consideration of the implications of this are looked at in the next section.

5.2.2 Paired data

The composition of the data set for the paired analysis will be slightly different to that for the unpaired analysis as the total number of subjects was less due to dropout at various stages of the data collection so giving a slightly different data set for the paired analysis. The two tables below show the descriptive data for the subjects who undertook setups on both the VERT and LINAC.

Table 5.3 Group descriptives (Paired data)

N	Age Mean and S.D.	Male to Female ratio	Cohort Ratios Years 1, 2, 3
44	24·3 (6·5)	13:31 (1:2·4)	4:24:16 (1:6:4)

Table 5.4 ITQ Results (Paired data)

	N	Total Mean & S.D.	Focus Mean & S.D.	Involvement Mean & S.D.	Games Mean & S.D.
Test data	44	74·6 (16·6)	31·3 (5·9)	28·0 (9·2)	5·8 (3·5)
Normative data	152	76·7 (13·6)	40·3 (6·1)	26·5 (7·2)	6·2 (3·2)
Test of difference		t = 0·8567, p = 0·3927	t = 8·6808, p < 0·0001	t = 1·1397, p = 0·2558	t = 0·7148, p = 0·4756

Normative data taken from Witmer & Singer, (1998)

The pattern of similarity and difference between the tests data ITQ scores and the normative data identified in Table 5.2 was again repeated with the paired data above in Table 5.4. Both this study and the study by Witmer and Singer used students, although Witmer and Singer failed to describe the age bracket of the students that formed their sample. A number of differences did exist between the two studies that might explain the variation; firstly the country the studies were undertaken in is different and secondly there are differences in the gender makeup of the two studies. This study contained a far higher proportion of females compared to Witmer and Singer's population that had more males than females by a ratio of approximately 1·5:1. This rationale for the difference might be supported by looking at a study undertaken by Murray, Fox & Pettifer (2007) in the UK on 40 females and 22 males. Although they didn't report the subscale scores the mean total score of 68·23 (S.D. 13·18) was considerably lower than that of Witmer and Singer's total score, which might imply a naturally wide normal variation of scores between countries and genders for this instrument.

5.3 Inferential Statistics

The following sections in this chapter deal with the six aims that are being dealt with through quantitative analysis. Each aim has been given its own section.

5.4 Aim 1. To discover if Immersive tendency predicts the feeling of presence in both simulators.

$H0_1$ There is no relationship between immersive tendency and presence.

In order to discover if immersive tendency can predict presence, correlations were undertaken between the ITQ scores and Presence Questionnaire (PQ) scores. Firstly scores that were available for all subjects (n=50) were correlated, then correlations were undertaken again looking at the data from the LINAC and VERT simulations separately.

Table 5.5 Correlations between ITQ and PQ scores: All students.

PQ ITQ	Involvement	Sensory fidelity	Adaptation Immersion	Interface quality	Total
Total	R = 0·136, p = 0·346	R = 0·111, p = 0·433	R = 0·107, p = 0·458	R = -0·106, p = 0·464	R = 0·108, p = 0·455
Focus	R = 0·129, p = 0·373	R = 0·106, p = 0·464	R = 0·175, p = 0·223	R = -0·023, p = 0·872	R = 0·132, p = 0·361
Involvement	R = 0·039, p = 0·786	R=0·071, p = 0·622	R = -0·070, p = 0·627	R = -0·175, p = 0·224	R = -0·003, p = 0·983
Games	R = 0·306, p = 0·031	R=0·231, p = 0·106	R = 0·316, p = 0·025	R = 0·124, p = 0·392	R = 0·301, p = 0·034

Table 5.5 above shows that there is no relationship between ITQ and PQ scores except for the construct Gaming on the ITQ questionnaire and three presence scores for the constructs of involvement, adaptation/immersion and the total score. When looking at the individual simulations it can be seen from Table 5.6 that this relationship exists only on the VERT simulation and not on the LINAC simulation and a further relationship was identified for the VERT simulation between gaming and sensory fidelity.

Table 5.6 Correlations between ITQ and PQ scores: Unit dependent scores.

PQ / ITQ	Involvement	Sensory fidelity	Adaptation Immersion	Interface quality	Total
LINAC Total	R = 0·133, p = 0·555	R = 0·083, p = 0·715	R = 0·014, p = 0·952	R = -0·233, p = 0·296	R = 0·037, p = 0·870
VERT Total	R = 0·119, p = 0·546	R = 0·103, p = 0·602	R = 0·127, p = 0·521	R = -0·049, p = 0·805	R = 0·122, p = 0·537
LINAC Focus	R = -0·093, p = 0·682	R = 0·036, p = 0·874	R = -0·016, p = 0·943	R = -0·148, p = 0·510	R = -0·067, p = 0·767
VERT Focus	R = 0·246, p = 0·206	R = 0·092, p = 0·642	R = 0·259, p = 0·184	R = 0·046, p = 0·815	R = 0·235, p = 0·229
LINAC Involvement	R = 0·296, p = 0·181	R = 0·151, p = 0·501	R = -0·033, p = 0·882	R = -0·291, p = 0·189	R = 0·102, p = 0·651
VERT Involvement	R = -0·139, p = 0·480	R = 0·040, p = 0·840	R = -0·128, p = 0·518	R = -0·123, p = 0·543	R = -0·113, p = 0·568
LINAC Games	R = 0·172, p = 0·445	R = 0·030, p = 0·895	R = 0·128, p = 0·570	R = 0·072, p = 0·750	R = 0·132, p = 0·559
VERT Games	R = 0·460, p = 0·014	R = 0·397, p = 0·036	R = 0·446, p = 0·017	R = 0·119, p = 0·546	R = 0·487, p = 0·009

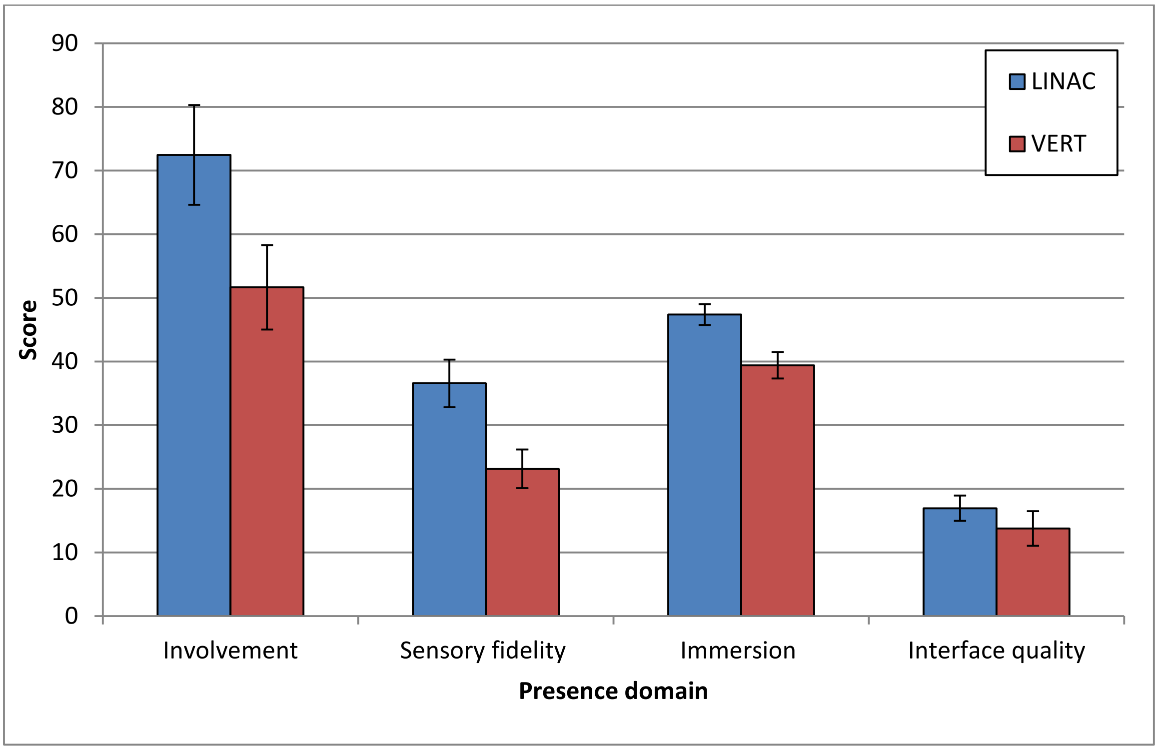
5.5 Aim 2. To assess if Presence will be higher on the LINAC simulation compared to the VERT simulation.

$H0_2$ There will be no difference in Presence scores between the two simulation methods.

After completing each series of setups the students were asked to complete the Singer Presence questionnaire in order to evaluate in which simulation the subjects who completed the questionnaire felt most presence. Results shown in Figure 5.1 from the four domains of the questionnaire show that subjects rated the LINAC setups consistently higher than the VERT setups and that the difference was significant in all four domains.

The mean difference in the total presence scores between the LINAC simulation and the VERT simulation was 44·1, the presence score being higher on the LINAC (173·0) than it was on VERT (128·9). The difference was statistically significant t = 6·462, p < 0·001 indicating that we can reject the null hypothesis and accept the alternate hypothesis that students experience more presence on the LINAC simulation compared to the VERT simulation. Almost identical results were observed in the paired test the overall mean score for presence again being significantly higher on the LINAC (173·8, 95% CI 166·7 to 180·1) compared to VERT (130·4, 95% CI 120·1 to 135·8), F = 91·93, p < 0·001. The four domains that make up this overall score can be seen below in Figure 5.1.

Figure 5.1 Presence scores for the VERT and LINAC simulations.



Mean score with 95% confidence intervals.

Further analysis of the sub-scales present (Table 5.7) within the questionnaire also identified a significantly higher level of involvement, sensory fidelity, immersion and interface quality for the LINAC compared with VERT.

Table 5.7 Simulation Presence scores.

PQ Scale	Unit	N	Mean score	S Dev	Significance
Involvement	LINAC	22	72·7	10·0	t = 6·175, p < 0·0001
	VERT	30	53·9	11·5	
Sensory fidelity	LINAC	22	36·8	4·5	t = -8·169, p < 0·0001
	VERT	30	23·8	6·4	
Adaptation/ Immersion	LINAC	22	47·1	5·7	t(d) = -4·117, p = 0·0001
	VERT	30	38·6	9·2	
Interface quality	LINAC	22	17·2	4·1	t(d) = -3·047, p = 0·0045
	VERT	30	14·1	2·7	

t(d) = Modified t-test, Welch's test utilised due to unequal variance within the sample.

5.6 Aim 3. What student characteristics moderate the presence scores?

$H0_3$ Age and gender will not affect the presence score.

The overall PQ score and the scores for the domains was tested firstly using a t-test to assess if differences existed in different genders. As can be seen from Tables 5.8 and 5.9 males scored higher compared to females for presence on both the LINAC and VERT simulations in all domains; although significance was not reached on the LINAC for interface quality and on VERT for sensory fidelity. Males therefore felt more presence in both the VERT and the LINAC simulations compared to females.

Table 5.8 The effect of gender on Presence score (LINAC).

Gender	Involvement	Sensory Fidelity	Adaption/ Immersion	Interface Quality	Total Score
Males	77·9	39·0	51·8	18·2	186·8
Females	70·6	35·7	45·2	15·6	168·1
Difference	7.2	3.3	6.5	1.6	18.7
Sig	t(d) = 3·246, p = 0·002	t(d)= 2·803, p = 0·008	t(d) = 5·084, p < 0·001	t = 1·201 p = 0·237	t(d) = 2·894, p < 0·001

t(d) = Modified t-test, Welch's test utilised due to unequal variance within the sample.

Table 5.9 The effect of gender on Presence score (VERT).

Gender	Involvement	Sensory Fidelity	Adaption/ Immersion	Interface Quality	Total Score
Males	58·4	24.8	44·0	15·5	142·6
Females	50·3	23·3	37·9	13·2	124·7
Difference	8·1	1.5	6.1	2.3	17.9
Sig	t = 2·355, p = 0·023	t = 0·792, p = 0·128	t = 2·495, p = 0·016	t = 2·818 p = 0·007	t = 2·567, p = 0·013

There was no relationship between age and presence score noted on both simulation, Tables 5.10 and 5.11. A number of values approached significance for the LINAC simulation. Interface quality and age for all subjects and also for females showed medium positive relationship. In males, involvement, sensory fidelity and total PQ score all had a strong positive association with age.

Table 5.10 Correlations between Age and PQ scores: Gender dependent scores. (LINAC)

PQ ITQ	Involvement	Sensory Fidelity	Adaptation Immersion	Interface quality	Total
Age-All	R = 0·108, p = 0·489	R = 0·187, P = 0·229	R = 0·168, p = 0·281	R = 0·309, p = 0·053	R = 0·203, p = 0·192
Age-Female	R = -0·023, p = 0·904	R = 0·097, P = 0·609	R = 0·097, p = 0·611	R = 0·347, p = 0·061	R = 0·106, p = 0·576
Age-Male	R = 0·534, p = 0·060	R = 0·537, P = 0·058	R = 0·165, p = 0·590	R = 0·105, p = 0·733	R = 0·546, p = 0·054

Table 5.11 Correlations between Age and PQ scores: Gender dependent scores. (VERT)

PQ ITQ	Involvement	Sensory Fidelity	Adaptation Immersion	Interface quality	Total
Age-All	R = 0·138, p = 0·345	R = 0·019, P = 0·896	R = 0·092, p = 0·529	R = 0·207, p = 0·154	R = 0·207, p = 0·380
Age-Female	R = -0·017, p = 0·922	R = -0·081, P = 0·643	R = -0·039, p = 0·825	R = 0·210, p = 0·225	R = -0·018, p = 0·919
Age-Male	R = 0·474, p = 0·087	R = 0·165, P = 0·573	R = 0·437, p = 0·118	R = 0·089, p = 0·762	R = 0·397, p = 0·159

5.7 Aim 4. Do the outcomes of the two simulated setups utilising different equipment agree with each other?

$H0_4$ There will be no difference in simulation setup parameters for the two simulation methods.

5.7.1 Time taken to perform the setups

The time subjects took to undertake the assessments did vary between the two simulation methods, but not significantly, see Table 5.12. The total time needed to undertake both electron setups was longer for the LINAC based simulation by just over a minute. This small difference appears to arise mainly from the time taken to undertake the second electron setup, the mean time taken for the first electron setup being very similar between the two simulation methods. Although there is a difference in the mean time taken to undertake each electron setup the difference does not reach significance.

Table 5.12 Time taken to undertake the electron setups.

	Setup 1		**Setup 2**		**Total time**	
Unit	LINAC	VERT	LINAC	VERT	LINAC	VERT
N	22	30	22	28	22	28
Mean time (mins)	6·38	6·28	5·0	3·7	11·4	9·9
S Dev	3·48	2·97	2·47	2·36	4·95	4·76
Sig	t = 0·014, p = 0·917		t = 1·882, p = 0·066		t = -1·07, p = 0·287	

5.7.2 Difference between setups

When comparing student competency on each simulation two possible routes of analysis are available. The data can be looked at as both ranked data by looking at the score of all the sub components that make up the competency. Alternatively it can be looked at as a dichotomous variable based on whether setup was acceptable or not. Dichotomous scores were considered first.

Table 5.13 overleaf looks at the setup competency as a dichotomous variable and at the individual components that make up the overall competency, the main difference between the VERT and LINAC setups being the fit of the light beam in the medial/lateral direction. In setup 1 there were also significant differences between the setups for the fit of the light beam in the superior/inferior direction and the correct setting of the treatment distance.

In all cases except for the dose variance in setup 2 the LINAC performance was better than that of the VERT performance as can be seen in Figure 5.13.

Table 5.13 Comparison of setup competency components.

	Setup 1		Setup 2	
	Fisher's Exact test	**Significance**	**Fisher's Exact test**	**Significance**
Overall Acceptable setup	6·83	0·017	1·65	0·643
Rotation fit	1·68	0·194	2·529	0·480
Superior/Inferior fit	7·06	0·003	4·676	0·108
Medial/Lateral fit	9·30	<0·001	13·762	<0·001
Machine hit the patient	0·144	1·000	0·748	1·000
Treatment distance	11·6	<0·001	3·771	0·228
<5% Dose Variation	0·400	0·4485	1·679	0·778

Figure 5.2 below shows the performance of the subjects for the above analysis, the figure showing the percentage of students that performed the part of the setup correctly. The students performed better on the LINAC for all measures except dose variance which was performed better on VERT for the second setup.

Figure 5.2 Comparison of setup components.

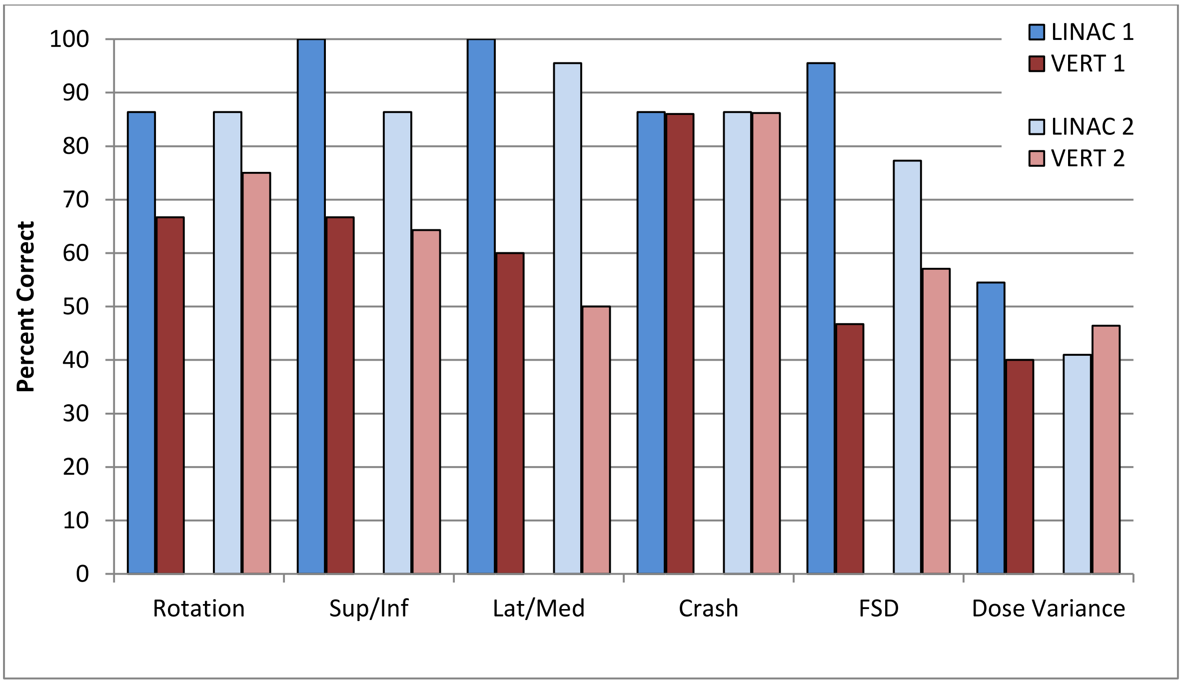
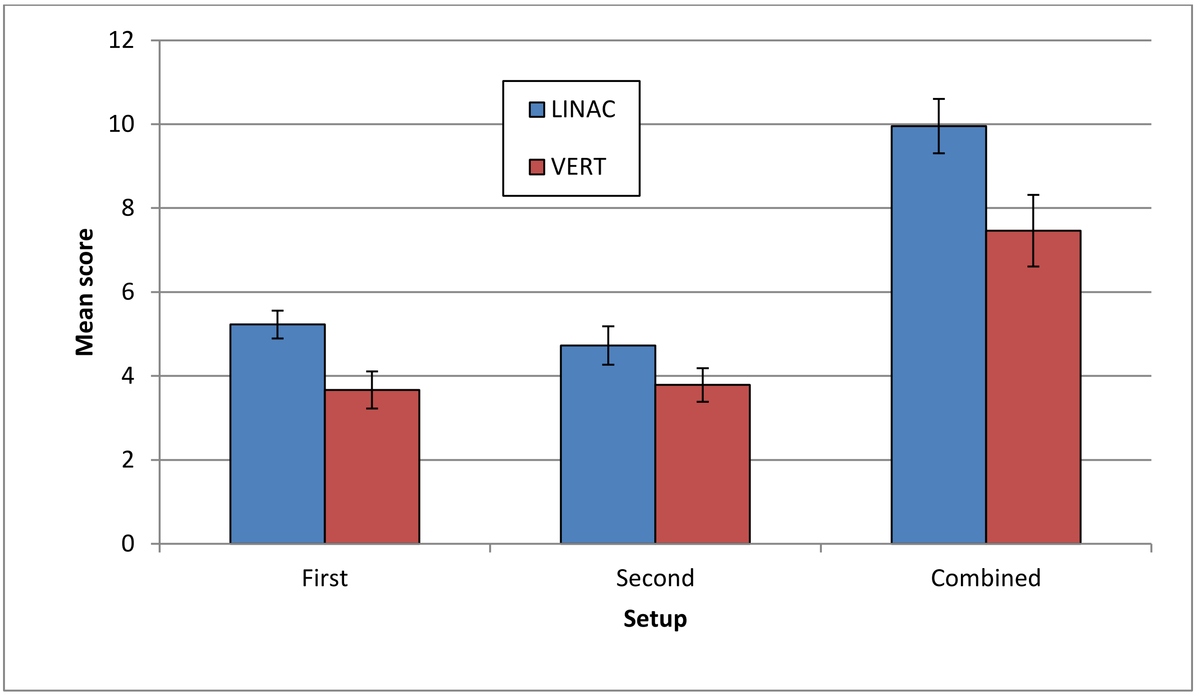


Figure 5.3 shows that overall, students performed better on the LINAC for both setups. Scores were derived from scoring each setup using the six competency measures seen in Table 5.6 so giving each participant a score between 0 and 6 for each setup. The difference

in all scores between VERT and the LINAC simulations were significant, U = 96·0, p < 0·0001, U = 172·0, p = 0·0062 and total score of U = 110·5, p < 0·0001.

Figure 5.3 Competency score for setups.



Mean score with 95% confidence intervals

5.8 Paired analysis

The intention in this section is to sequentially repeat the steps in the unpaired analysis in the previous section for ease of comparison later. Additional tests were also performed in this section due to the nature of the data and to give a full analysis.

5.8.1 Time taken to perform the setups

The paired analyses are presented in the same order as the unpaired analyses starting with the time taken to undertake both setups. The results of the paired analyses are shown in Table 5.14.

Table 5.14 Time taken to undertake the electron setups.

	Setup 1		Setup 2		Total time	
Unit	LINAC	VERT	LINAC	VERT	LINAC	VERT
N	44		40		40	
Mean time (mins)	5·82	5·39	5·46	3·33	11·30	8·59
S Dev	2·82	2·43	3·08	1·99	4.88	4.06
Sig	F = 0·95, p = 0·335		F = 28·34, p < 0·001		F = 16·79, p < 0·001	

No order effect was seen for the second setup and for the total time (F = 0·10, p = 0·751, and F = 1·21, p = 0·278 respectively), but was seen for setup 1, F = 4·43, p = 0·041. This indicates that in the case of setup 1 there was an order interaction which indicates that the sequence of the setup did affect the time taken. The results for setup 1 and setup 2 are broadly in line with the unpaired analysis in that in both instances the students spent more time on the LINAC setup than the identical VERT setup although significance was only achieved in setup 2. Although both analyses showed that the total time took longer on the LINAC compared to VERT only the paired analysis reached significance. In hypothesis testing variability between individuals within the sample decreases the power of a test. Because of this the paired test has an advantage over the unpaired test in that the power of the test is increased as each person is acting as their own control so reducing the effects of any confounding variables; however this type of test is prone to experience effects, (Privitera, 2011), which is why it was important to look at order effect which was seen for setup 1.

5.8.2 Difference between setups

Three tests are commonly put forward for the analysis of binary data from crossover trials, the McNemar, Mainland-Gart and the Prescott test. Senn (2002) recommends the McNemar's test for an analysis of binomial data when ignoring the effect of period or when no crossover effect exists. However, both the Mainland-Gart and Prescott tests into account the effect of the period i.e. the sequence the events took place in, and if there is a period effect the test are invalid. The Mainland-Gart test is more robust than the Prescott test as it does not depend on the subjects being randomised into groups, however the analysis only considers differences in periods which can mean that the analysis is being performed on only a small proportion of patients in the study as all tied data is not considered. In this situation Pictor, (2003) argues that the Prescott test is more advantageous to use than the Mainland-Gart test. As much of the data produced was tied and the importance of considering the period in the analysis it was decided to utilise the Prescott's test $X^2(P)$ for the analysis of the binary data.

Binary outcomes were analysed using the Prescott test obtained from Jones & Kenward (1989) which is shown below. The formula produces a chi squared approximation that corrects for continuity and looks for differences in treatments whilst considering the order the outcomes were produced.

$$X^2(P) = \frac{[|n12 - n13)n.. - (n2 - n3)n1| - \frac{1}{2}n..]^2}{\left[\frac{n1\ n2\ [(n2+n3)n.. - (n2-n3)]^2}{n..}\right]}$$

The formula uses symbols derived from the following contingency table.

Table 5.15 Prescott's contingency table.

Sequence	Preference for First period	No preference for either period	Preference for Second period	Total
VERT/LINAC	n12	n11 + n14	n13	n1
LINAC /VERT	n22	n21 + n24	n23	n2
Total	n2	n1 + n4	n3	n..

This test is not part of any statistical package and therefore calculations were undertaken in Excel. In order to ensure that the formula stated above was correctly written for Excel prior to looking at the data from the study values from two existing publications were first entered into the Excel spreadsheet and the derived answers confirmed against the published answers.

Table 5.16 Prescott's test for setup competence components. (Paired data)

N=44	Setup 1		Setup 2	
	$X^2(P)$	Significance	$X^2(P)$	Significance
Rotation fit	5·49	0·0191	6·69	0·0097
Superior/Inferior fit	17·04	<0·0001	6·74	0·0095
Medial/Lateral fit	15·13	0·0001	13·46	0·0002
Machine hit the patient	0·11	0·7400	0·10	0·7516
F.S.D.	11·25	0·0008	2·12	0·1455
<5% Dose Variation	1·84	0·1750	0·002	0·9902
Overall	12·51	0·0004	4·61	0·0319

The competency variables were also tested to see if there was a period effect that could confound any observed treatment effect, again following the guidelines identified by Jones & Kenward (1989). No significant period effects were observed, Table 5.17 indicating that any difference between the two simulation setups was because of the setups rather than the order they were performed in.

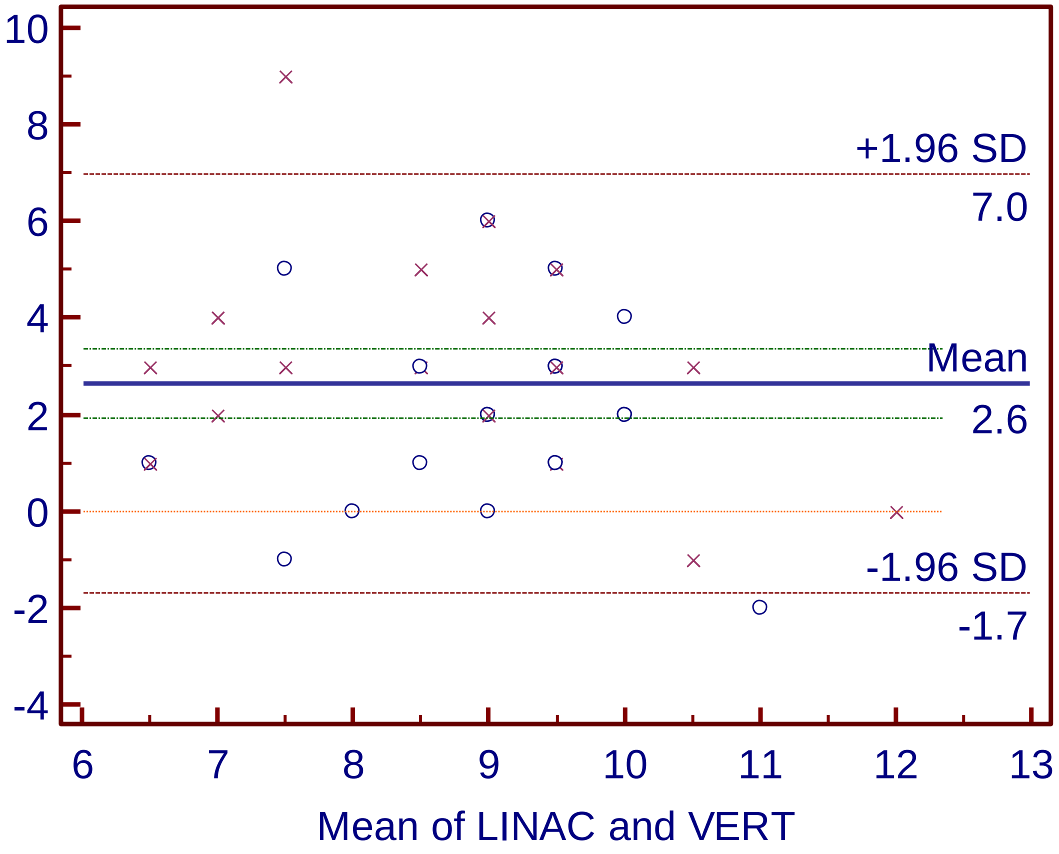
Table 5.17 Prescott's test results for period effects. (Paired data)

N=44	Setup 1		Setup 2	
	$X^2(P)$	Significance	$X^2(P)$	Significance
Rotation fit	1·98	0·1594	0·66	0·4153
Superior/Inferior fit	0·02	0·8986	0·01	0·9105
Medial/Lateral fit	0·01	0·9078	3·45	0·0630
Machine hit the patient	0·11	0·7400	0·10	0·7516
F.S.D.	0·79	0·3736	1·06	0·3043
<5% Dose Variation	3·35	0·0673	<0·001	0·9902
Overall	0·03	0·8542	1·24	0·2659

The focus on the statistical tests to date has been on looking for a difference between groups, however Bland & Altman (1986) suggested a plot to look at similarities between paired measures which is shown in Figure 5.4. The graph is shown as an interesting counterpoint to the statistical analysis comparing performance of students on the two simulations.

If students performed equally well on both simulations we would expect the data points to be around 0 as shown by the dotted orange line. The graph shows that there is a clear movement towards a positive score irrespective of how well they performed, or which simulation they performed first; the mean LINAC score being 2·6 higher than the VERT mean score. Of the students undertaking the study only 3 students performed equally well on both simulations (the three data points on the orange line), two of whom did the LINAC simulation first and one who did the VERT simulation first.

Figure 5.4 Bland Altman plot of overall competency scores



Circles =LINAC/VERT, Crosses=VERT/LINAC

5.9 Aim 5. What factors moderate the simulation scores?

$H0_5$ Cohort, gender and immersive tendency will not affect the simulation score.

When looking at gender performance (Table 5.18) females did take longer to undertake the setups on 3 of the four setups undertaken, the only setup females were faster on was the second VERT one and then only by a mean of 0·08 minutes, (5 seconds). Although males were quicker on 3 of the setups and overall by 2·84 minutes (2 minutes 50 seconds) on the LINAC and 0·51 minutes (31 seconds) on VERT no significant difference was observed on any timing. When each gender's time was considered for each simulation it could be seen that for the first simulation males were faster on the LINAC for both of the simulations. For the second setup males were again faster on the LINAC setup whereas females were faster on the VERT setup, females by a significant margin, therefore most of the difference observed in Table 5.12 for the second setup comes from the females, not the males. This difference carries over into the overall time which again was significantly different for the population, females again making up most of the variation.

Table 5.18 The effect of gender on setup time.

		LINAC time (min)	VERT time (min)	Difference (min)	Significance
Setup 1	Males	4·78	5·33	-0·55	t = 0·49, p = 0·622
	Females	6·32	5·92	0·40	t = -0·52, p = 0·608
	Difference (min)	1·54	0·59		
	Significance	t = 1·61, p = 0·115	t = -0·70, p = 0·482		
Setup 2	Males	4·50	3·51	0·99	t = -0·91, p = 0·371
	Females	5·80	3·43	2·37	t = -3·97, $p < 0·001$
	Difference (min)	1·30	0·08		
	Significance	t = -1·85, p = 0·072	t = 0·12, p = 0·907		
Total time	Males	9·28	8·84	0·44	t = -0·22, p = 0·827
	Females	12·12	9·36	2·77	t = -2·78, p = 0·007
	Difference (min)	2·84	0·51		
	Significance	t = -1·32, p = 0·193	t = -0·12, p = 0·907		

When looking at performance in terms of the competency scores it can be seen that males outperformed females on the first LINAC setup, but females outperformed males on the equivalent VERT setup. This pattern was reversed for setup 2 with females outperforming males on the LINAC setup and males outperforming females in VERT. Only the females outperforming the males on the second LINAC setup reached significance.

When looking at each genders performance on both VERT and the LINAC simulations it can be seen that although there is some variation, both males and females performed significantly better on the LINAC simulation compared to the VERT simulation.

Table 5.19 The effect of gender on setup score.

		LINAC Score	VERT Score	Difference	Significance
Setup 1 score	Males	5·54	3·57	1·97	W = 78·00, p < 0·002
	Females	5·17	3·97	1·20	W = 366·00, p < 0·001
	Difference	0·37	-0·04		
	Significance	U = 2·79, p = 0·181	U = 0·54, p = 0·724		
Setup 2 score	Males	4·42	4·00	0·42	W = 49·00, p = 0·026
	Females	4·69	3·77	0·92	W = 228·00, p = 0·005
	Difference	-0·27	0·23		
	Significance	U = 6·23, p = 0·033	U = 0·01, p = 0·768		
Total score	Males	9·96	7·57	2·39	W = 78·00, p = 0·002
	Females	9·86	7·74	2·12	W = 348·00, p < 0·001
	Difference	0·10	-0·17		
	Significance	U = 1·10, p = 0·476	U = 0·26, p = 0·863		

The data was also looked at to see if the different cohorts obtained different scores on each type of simulation. Table 5.20 indicates no significant difference in competency score between the cohorts undertaking the electron setups, although it is worth noting that the LINAC score does increase with progressive cohort years whereas the VERT score does not.

Table 5.20 Competency score by Cohort.

	LINAC (Score)			VERT (Score)		
	n	**Rank**	**Mean**	**n**	**Rank**	**Mean**
Year 1	3	8·83	9·7	4	18·00	8·5
Year 2	11	11·09	10·0	16	11·97	6·8
Year 3	8	13·06	10·4	8	17·81	8·3
Significance	F = 0·496, p = 0·617			F = 1·743, p = 0·196		

The relationship between presence as measured by the PQ questionnaire and performance was also investigated. Correlations were performed between the various sub-scales of the PQ questionnaire, ITQ questionnaire and the overall performance score on both the VERT and the LINAC simulations. The observed correlation coefficients with overall performance could at best be described as being weak, the strongest being a presence sub-scale "sensory fidelity" on the LINAC with a correlation coefficient ρ of 0·29. No correlation coefficient approached significance.

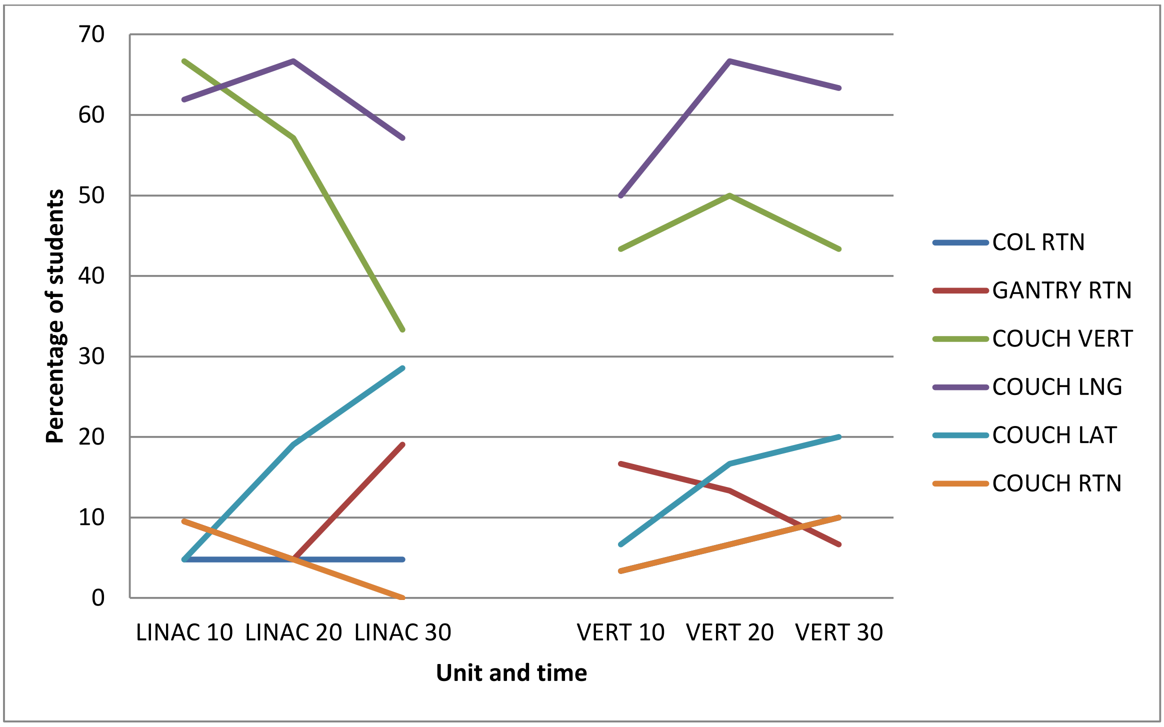
5.10 Aim 6. Do participants utilise the same cognitive process on both systems?

5.10.1 Machine movement at the start and end of the setup.

As seen earlier in Table 5.14 the time taken to undertake the setups was very different for individual students and as observations were only taken at 10 second intervals comparison of the students' setups between VERT and the LINAC was problematic. It was therefore decided to look at student activity during the first 30 and final 30 seconds of both setups. Data for the first setup are shown in Figures 5.5 and 5.6.

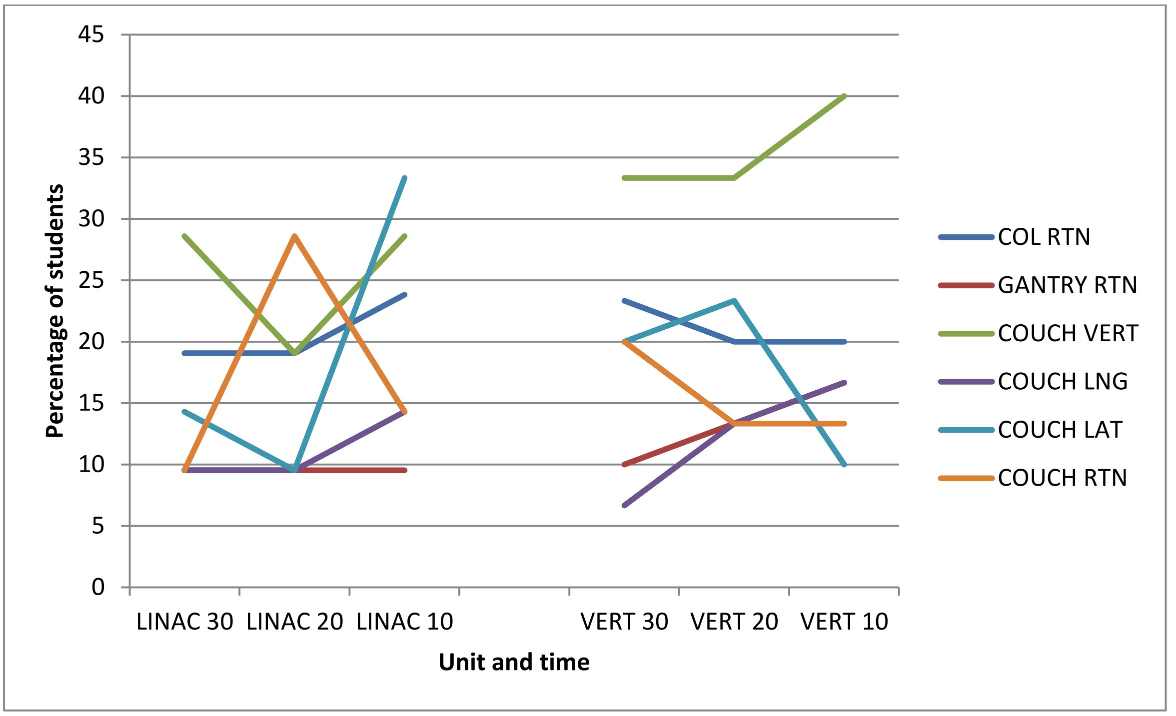
Similarities can be observed in the movement patterns especially at the start of the setup. In Figure 5.5 it is noticeable that both the couch long and couch vertical movements are the most commonly performed movements on both simulations during the first 30 seconds. However, there are also noticeable differences. It can also be seen that the use of the gantry rotation and couch rotation movements appears to be reversed, on the LINAC use of gantry motion increases over the first 30 seconds whilst couch rotation use it at its highest in the first 10 seconds and then decreases whereas on VERT gantry rotation is most predominantly used in the first 10 seconds after which its use decreases whereas couch rotation use slowly increases and is most used during the final 10 seconds.

Figure 5.5 Machine parameter movements during the first 30 seconds. 1st setup.



Movements during the last 30 seconds of the setup (Figures 5.6) were more varied between the units. Two noticeable differences between the student's use of the two units was the greater use of the couch vertical movement on the VERT and the increase in the number of students using the couch lateral movement on the LINAC in the final 10 seconds compared to the preceding time period which is the opposite of what tends to happen on VERT which shows a decrease in use of the couch lateral movement in the final 10 seconds of the setup.

Figure 5.6 Machine parameter movements during the final 30 seconds. 1st setup.



The analysis was re-performed for the second electron setup, the results of which can be seen in the following two graphs. The pattern is noticeably different to the first setup as the start point for the second setup was the end point for the first setup.

Figure 5.7 Machine parameter movements during the first 30 seconds. 2nd setup.

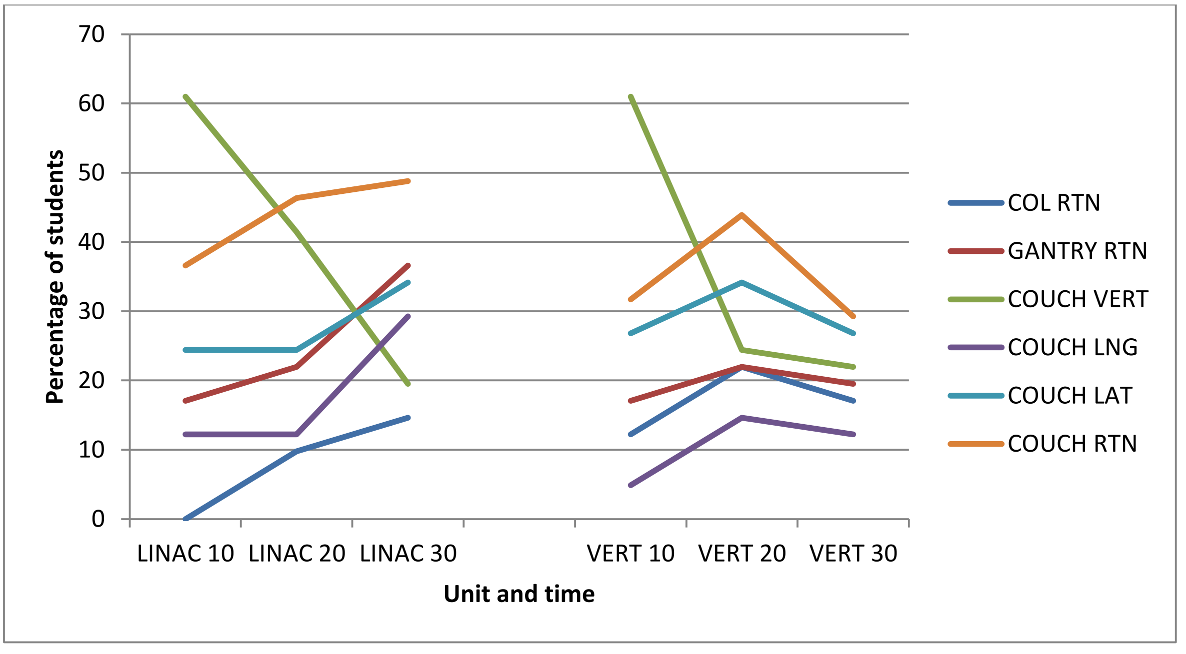
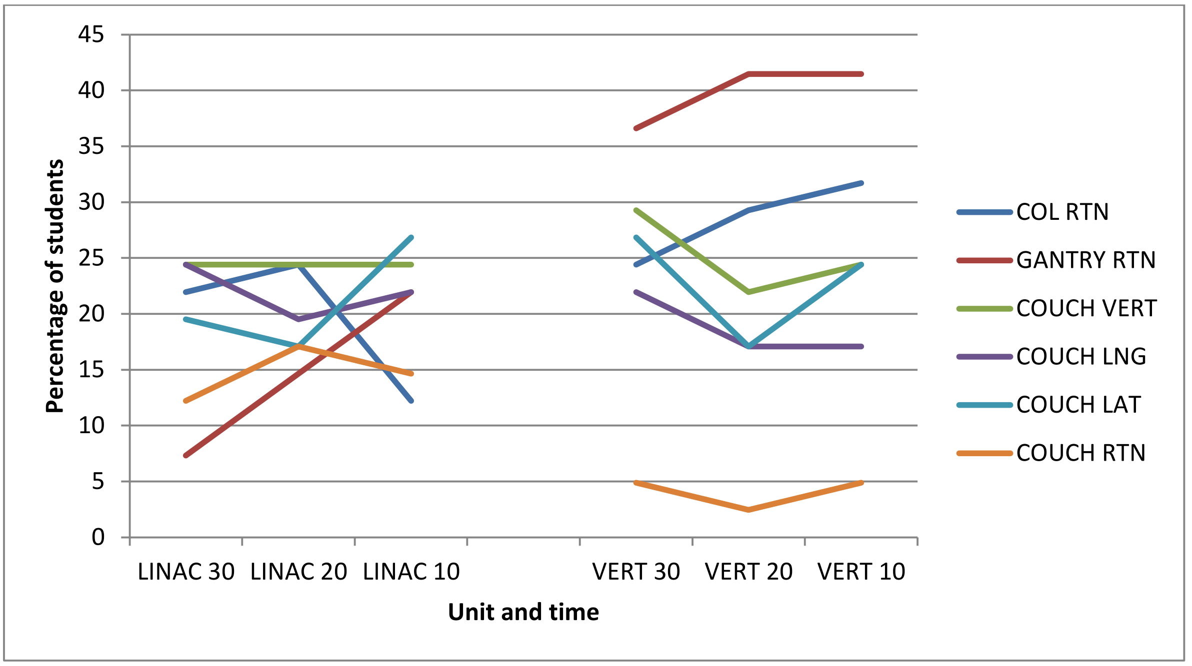


Figure 5.8 Machine parameter movements during the last 30 seconds. 2nd setup.



As with the first setup there appears to be more similarity in the movements during the first 30 seconds of the machine's movement compared to the final 30 seconds of movement. Both at the start and end of the second setup there appears to be a lot less use of the couch rotation movement on the VERT compared to the LINAC and in the final 30 seconds a lot more gantry movement on VERT. Although these findings cannot be linked directly to the competency results the differences in use between the two units observed in the final 30 seconds might in some way be associated with the differences observed in some of the earlier results such as time and competency score.

5.10.2 Student inactivity during the setup.

Periods of inactivity were also noted from the video evidence. This data was generated from the observations taken every 10 seconds and therefore does not represent a 10 second period of inactivity, but rather no activity during the 10 second video period.

Figure 5.9 User inactivity.

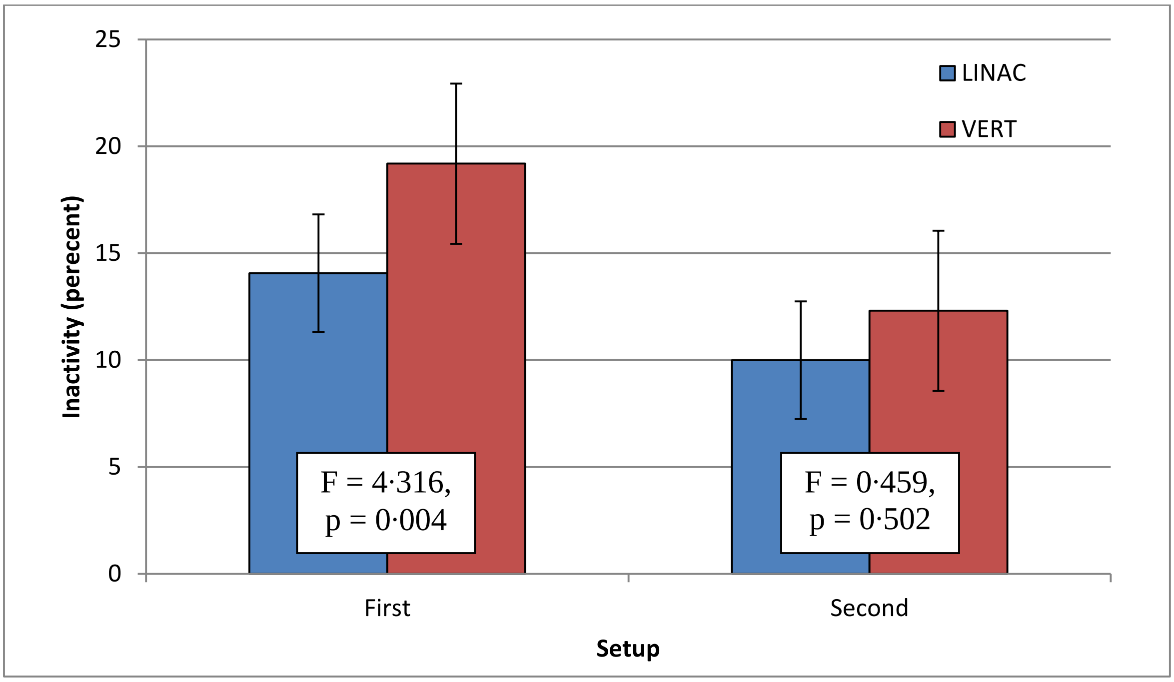
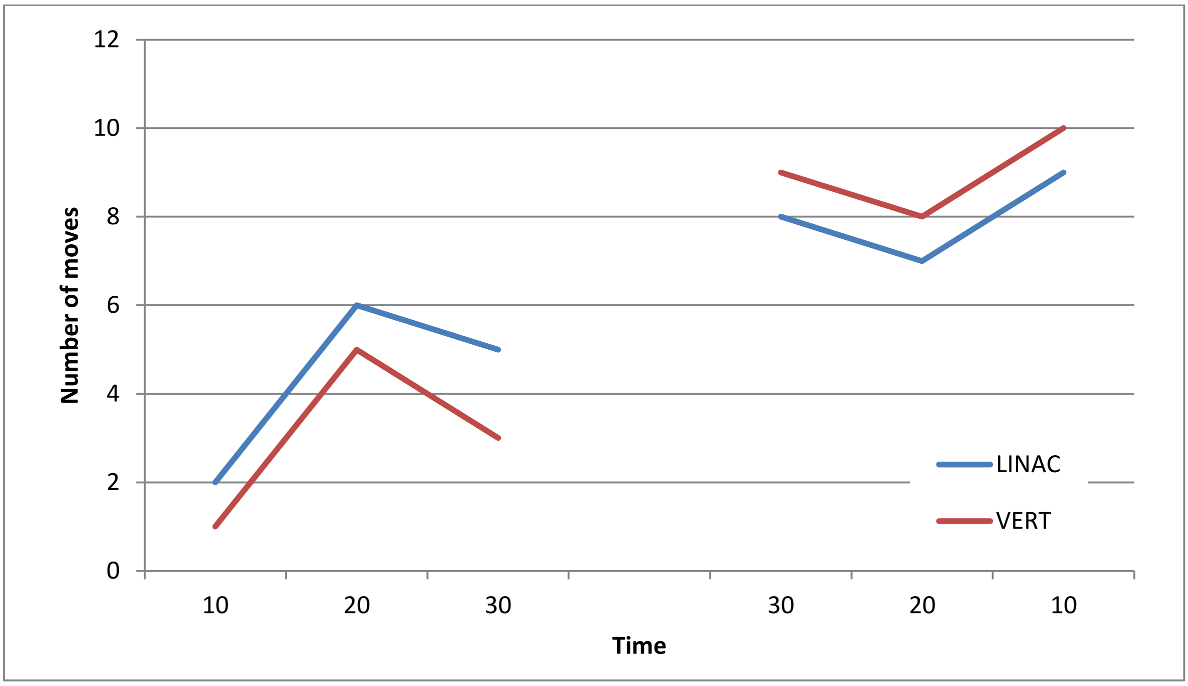


Figure 5.9, above indicates that students had more periods of inactivity on VERT than they did in the LINAC simulation. The difference in periods of inactivity reached significance only for the first setup where the difference in activity between VERT and LINAC simulations was 5·12.

No order interactions were observed for either setup, p = 0·107 and p = 0·881 respectively. The issue of which point in time students were inactive during suffers from the same issue identified in the first analysis within this section, namely that the time taken to undertake the setups varied considerably so a linear analysis becomes problematical. The following plot shows the number of movements in the first and final 30 seconds in both of the electron setups. The difference in inactivity identified in Figure 5.9 must occur during the setup as Figure 5.10 shows a marked resemblance at both the start and end of the setups.

Figure 5.10 Periods of user inactivity.



Because inactivity contributes to the time it takes to undertake the setup, correlations were then performed between the time for each setup and the periods of inactivity. Table 5.21 shows that three significant correlations exist and that the strongest association is seen for the two LINAC setups. Graphs showing these relationships can be found in Appendix X.

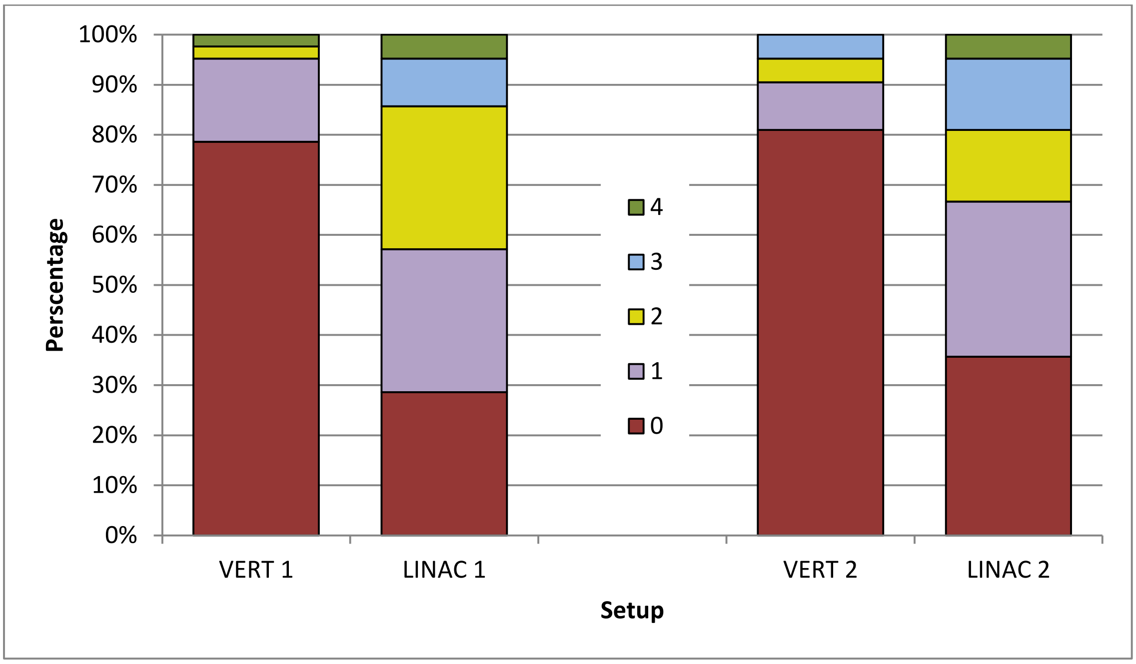
Table 5.21 Correlations between inactivity and time to undertake the setup.

	Amount of Inactivity			
	LINAC Setup 1	VERT Setup 1	LINAC Setup 2	VERT Setup 2
Time to setup	R = 0·372 p = 0·013	R = 0·123 p = 0·395	R = 0·326 p = 0·038	R = 0·369 p = 0·011

5.10.3 Multiple operations.

The observer present also recorded the number of times when the operator moved more than one treatment parameter at once. The average number of multiple operations on VERT was 0·31 for the first setup and 0·33 for the second setup indicating that most subjects did not do any double moves. On the LINAC simulation the average number of multiple movements was 1·33 and 1·12 respectively. Figure 5.11 clearly shows the similarity in the number of multiple moves in the two setups within each simulation, both vertical bars having similar patterns. It also identifies clear differences between the two different simulations, with approximately 20% of the students in both of the VERT simulations using multiple machine movements compared with between 72% and 64% on the LINAC.

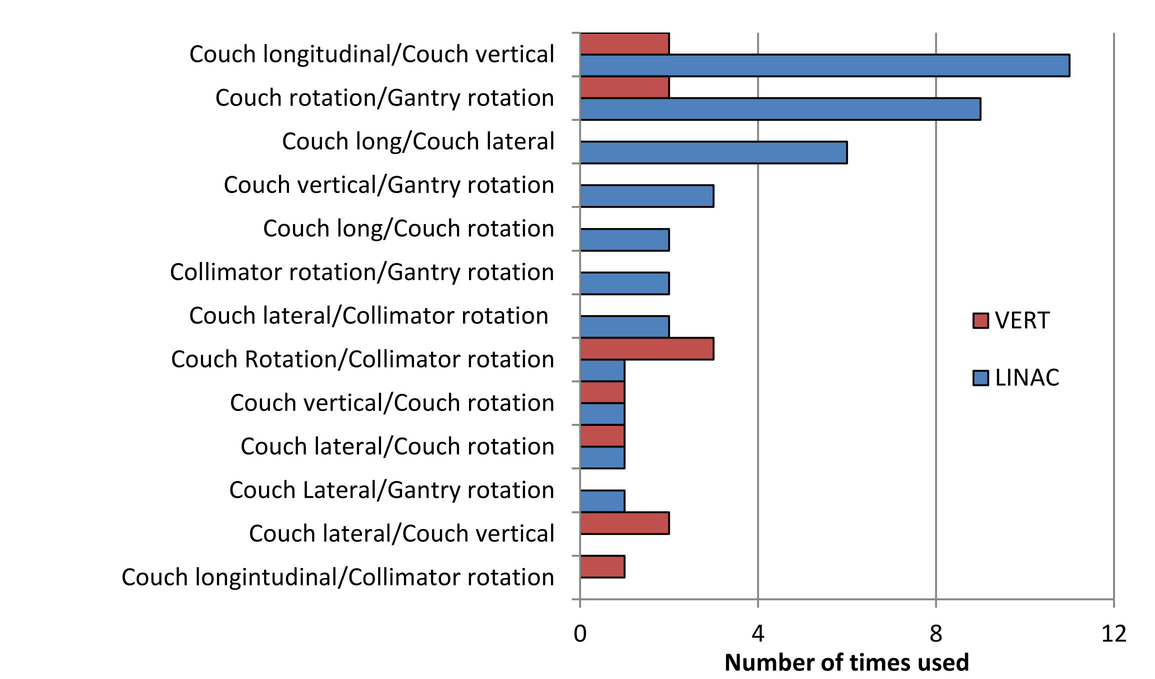
Figure 5.11 Students utilising multiple movements.



The differences observed between the two units on the use of multiple operations were both significant F = 27·352, p < 0·001 and F = 22·544, p < 0·001. There was no order effect for the first setup, p = 0·547, but one did exist for the second setup, p = 0·019 and because of this the analysis for the second setup was rerun using the unpaired data giving the following result, t(d) = 2·150, p = 0·039 confirming a statistical difference between the second setups on VERT and the LINAC.

Figure 5.12 overleaf shows the types and frequency of paired movements used. Only paired movements are shown for ease of interpretation and this type of movement made up the vast majority of the multiple movements used. The most common paired movement utilised by students was the couch longitudinal and couch vertical movement.

Figure 5.12 Paired movements used.



As with the analysis for inactivity the movement data was then analysed to see if it was related to the time taken, Table 5.22. The table indicates that only one significant correlation was observed between the number of double moves in the first LINAC setup and time taken and that this correlation was negative.

Table 5.22 Correlations between number of double moves and time to undertake the setup.

	Number of double moves			
	LINAC Setup 1	VERT Setup 1	LINAC Setup 2	VERT Setup 2
Time to setup	R = - 0·372 p = 0·013	R = -0·188 p = 0·192	R = 0·186 p = 0·239	R = 0·077 p =0·601

Chapter 6 - Qualitative Analysis

One of the aims of the study was to consider what students think about the appropriateness of both VERT and a LINAC based simulation as an assessment tool. The specific research questions asked were as follows:

1. How would you describe VERT as an aid/method to learning clinical skills?
2. Is there anything we can do to improve the VERT experience?
3. How would you describe VERT as a method of assessing competency for clinical work?
4. What suggestions would you make to improve the way VERT could be used to assess competency?

6.1 Identified Themes

This section will focus on the four major themes identified by the thematic analysis. The qualitative analysis identified a number of differences between the two simulation methods, with many students responding to the questions posed by comparing the two simulation methods highlighting strengths or weaknesses of one method compared to the other. There were four emergent themes which were derived from an inductive process which are detailed in the following sections. All quotes used in this section are coded by the use of two numbers, the first represents the focus group and the second the speaker. Codes were based on chronological order of the focus groups and the order the focus group participants first spoke.

6.2 Theme #1 Equipment use

In all the focus groups the majority of the subjects within the focus groups commented on practical issues of the two simulations. The major issue raised by the participants was that of the goggles/glasses worn by subjects in order to visualise the simulation in three dimensions in VERT.

Some students made broad-brush statements such as that made by student [2.2], "*I didn't like the glasses. Working without the glasses was best.*" Indicating both a dislike for the glasses and a preference to undertaking simulation activities without the glasses, but not stating what they disliked about the glasses.

Other students also picked up on the issue of the glasses supporting the dislike of the glasses, but also expanding on why they were not liked. Students commented mainly on the feel and fit of the glasses which they disliked, student [1.5] commenting that, "*It was a bit weird with the goggles on.*" Student [1.1] adding, "*They kept falling off every time you moved around.*" This feeling was supported in the third focus group when [3.4] stated, "*The glasses feel weird when you walk, when you have them on the glasses that's it, they felt weird and kept coming off.*"

The glasses falling off during the simulation was therefore a major issue, but not the only one experienced by students; students in focus group 2 also experienced issues with the wire connecting the glasses to the transmitter.

[2,4] "*I think I got tangled in them when they fell to the floor.*"
[2,1] "*Yeah, I got a bit tangled up as well.*"

The final practical issue identified with the LCS glasses was identified by students who normally wore prescription glasses. One student started the assessment in 3D but completed it in 2D, whilst another worked without her normal glasses as she didn't feel able to complete the assessment wearing both her normal glasses and the VERT glasses. The following section of transcript shows the student discussion relating to this aspect of the glasses use.

[1,4] "*Did you use 2D*?"
[1,5] "*Yeah,... because I couldn't wear 2 pairs of glasses*."
[1,3] "*I tried to do that as well, and I had issues because I tried to wear them and it was...*"
[1,2] "*Yeah, I didn't wear my glasses which didn't help, I just had the things*."
(Makes putting on glasses motion.)

The student's data that worked in 2D because she didn't feel as though she could wear the 3D glasses over her normal glasses was excluded from the quantitative analysis due to the lack of data generated working in 2D mode. Although most students did not wear prescription glasses, or if they did continued to wear them during the VERT simulation the effect of people undertaking the VERT assessment whist not wearing their glasses was not taken into account during the quantitative analysis as it was not recorded.

When students were asked "What suggestions would you make to improve the way VERT could be used to assess competency?" Students responded by again mentioning the glasses, suggesting that it was a major issue as they felt it affected their performance.

[4.5] "*Can we change the glasses?*"
[1,6] "*Was there another way that you could do that way with/without the glasses was there another way that we could have done it?"*

Another practical issue about the use of equipment during the simulation raised by all groups was that of the couch controls. The couch on the LINAC could be controlled by both the hand pendant and the controls on the couch itself whereas on VERT it could only be controlled by the hand pendant as the couch was virtual. There is a work around option offered by the VERT simulator that was not taken up in this study that of using an iPad with an app controller. Students had used these controls in the past as part of their training, the controls being simulated on an iPad which are activated by dragging your finger across the screen. This option was not made available for the study as students would have needed to hold both the iPad and the hand pendant, both of which need both hands to operate. Although not used as part of this study the iPad controller was commented on by student [3.2] who stated that,

"*You use the IPad thing for the other, for the couch controls that was a bit more weird I think. I think to get something a bit more realistic to the couch controls would be better.*"

The lack of couch controls on the VERT simulation was picked up as an issue by focus group 2. They looked at the use of couch controls twice, firstly as an advantage for the LINAC simulation, the presence of couch controls making a difference to them in undertaking the assessment and then later when looking at VERT as a negative commenting on their lack of presence.

[2.4] "*Yeah and you can move the couch, physically move the couch on the LINAC."*
(Chorus of yeahs, nods of heads)
[2,4] "*And you had to play with the controller and watch it go back and forth*."

[2,3] "*On VERT*!"
[2,4] "*Oh yes sorry, VERT*."
[2,5] "*Yes, we couldn't use the VERT couch like we could a LINAC*."

The benefit offered by couch controls on the LINAC simulation was also commented on by focus group 1.

[1,4] "*And you can use the couch as well*."
[1,6] "*Yeah, that makes a big difference*."

[1,1] "*I think it was part the couch controls as well I would have preferred to have used that….*"

Whereas focus groups 1 and 2 only stated a preference to having couch controls present and that it made a difference in the simulation, focus group 4 highlighted what this difference was to do with both the degree of control available and ease of use.

[4.5] "*Yeah, it was a lot more accurate as well because sometime when you used the pendant you meant just a flick, but it might be more than that. But with the couch you've got full control over it so you can adjust it even if it's like by a millimetre*."

[4.1] "*It's probably easier to use the couch controls and, and you can be closer whereas with the pendant you kinda got to be (looks up and down at pretend pendant in hand) I think it is just a bit easier to use the couch controls*."

Another issue raised relating to equipment by two of the focus groups was that of being able to use a ruler. Radiographers usually carry a ruler around with them to measure distances from the couch top, but these can also be used in electron setups to measure the standoff in the corners of the field. Some students in focus groups 1 and 2 commented that they used rulers, but that they could only be used on the LINAC simulation. Not all students recognised this and didn't use a ruler on either setup, but those that did said that being able to measure standoff on the LINAC helped with the setup.

[1,2] "*I used my rulers for the LINAC one*!"
[1,4] "*Yes, that helped*."
[1,1] "*Oh, that would have been the smart thing to do*."
[1,6] "*Yes, you were able to see more where you were and tweak it which you couldn't on VERT*."

6.3 Theme #2 Reality

Many students found the lack of reality in the simulations troubling. One aspect that detracted from reality for both simulations was the phantom itself. The first issue relating to the anthropomorphic phantom was that, [1.3] "*the patient doesn't move; it's an ideal patient.*" This was an issue for both the VERT and LINAC simulations as both used the same phantom, albeit one was virtual. The second issue with the phantom, again picked up by focus group 1 was the lack of reality caused by the type of phantom used, "*It's just a torso…. It would be better with a whole patient.*" [1.1]. This issue was also raised in other focus groups with some students referring to the phantom as a block, "*it's not a patient it's like a block.*" [4.1], "*I would just move the patient's chin whereas with the block it was a bit…*" [2.1]. The fact that the students sometimes referred to the phantom as a block rather than a phantom or patient further suggests that they deconstructed the phantom to nothing more than a shapeless piece of plastic rather than considering it as a real patient.

The patient not being real was also flagged up by focus group 3, with some students feeling that the simulation could be improved by making the patient more real in the participants' eyes by giving the patient a clinical history that could be read before the assessment started, "*Do you think we could maybe make the session more realistic? In the sense of the patient we have on the couch like it has a bit of background about the tumour, the patient that you're setting up.*" [3.4]. A further suggestion was to include live elements around the setup such as bringing a patient into the room, "*make it a bit more realistic, maybe even have someone that's acting as the patient kinda walking in, doing a chat,*" [3.4]

The effect on the students of not considering that the patient was real was that they took less care than in real life, students admitting that they were, "*a bit less considerate.*" [1.1]. Despite this being an issue with both simulations most students still only talked about this issue in relation to VERT.

"*I think because you have experienced being around a real live person you know, you've got to look at safety and when you've actually got a real live person there. You've actually got to look at the points; the things you have to follow when you are setting up a patient. When you've got VERT you're kind of not as vigilant when it comes to things like that.*" [2.2]. "*Yeah.*" [2.3].

"*do it on a LINAC and that's the real thing so it's better I would say, even if it is a dummy on the couch, or whatever, a fake patient.*" [1.1]

Another issue associated with a lack of reality that students felt affected their performance was the ability on VERT to pass through objects. This manifested itself in a number of ways

"*That's true, it wasn't the same on VERT, it felt strange too much seemed to move.*" ([2,1] and [2,3] nod) *"Yes, I think it affected what I did.*"
[2,1] "*Yes.*"
[2,4] "*Definitely.*"
[2,1] "*I think I walked through the couch!*" *(Laughter.)*
[2,4] "*Yeah, there was that. Or you would duck down to look and you would realise your head was in the couch.*"

The overall feeling was that VERT felt less realistic than the LINAC simulation. This was summed up by student [2.4] who stated that "*VERT was worse, the fact it wasn't real so you don't think about it.*" When challenged in the meeting that neither simulation was in fact "real" the response was that *"….one of them felt more real…because it was in the environment that you're more used to.*" [2.4], "*It was in a real life department.*" [2.2] "*You've got a physical machine in front of you, you can see movement and if it hits it, it's a bit more serious.*" [2.4] Having a real machine and physical phantom helped with students considering the LINAC simulation to be more real, not only that, but being in a radiotherapy department also helped with the reality of the LINAC simulation.

6.3.1 Reality: Environment and senses

Reality is a matter of perception and is not only affected by the objects you are working with such as the machine and patient, but the whole world in which the user is placed and their senses. Wider issues related to the environment and the students' perception of it also impacted on students' sense of reality. Participants expressed a number of issues related to the environment with participants contrasting the environments of both the LINAC and VERT, a number of students commenting that VERT, "*….didn't feel like a treatment room*" [3.1]. When asked why the reply was that "*It just wasn't the same as real life…. It felt different.. hospitals' have their own smell and (Laughter)"* [3.1].

Although other students laughed at this comment they did agree with the sentiment adding that, "*It didn't feel like a hospital.*" [3.3].

This issue with VERT was also picked up by some students in focus group 1 who stated that, "*…. the room at City was a bit like a computer room kinda thing was, erm, the way the room actually felt and everything didn't feel like* a*….*" [1.1], "*Like a bunker.*" [1.6]. Again students were acknowledging that the VERT room did not make them feel as though they were in a treatment room and in this case the sense in question was not smell as with focus group 3, but sound, "*It was all weird on VERT, sound, everything.*" [1.3], "*I think definitely the noise, the sound of that, of the machine is very strange.*" [1.6].

This view of the sound on VERT being wrong was commented on by members of focus groups 2 and 4. Student [4.1] commented that "*… the sound on the VERT sounds a little bit, fake. Like it's not the sound of the couch and whatever but it sounds a bit like…it doesn't sound quite right so I think that affects it.*" Two students in another focus group, [2.2] and [2.4] did acknowledge differences in sounds between the two simulations commenting that VERT was "quieter" and "different". However unlike students in focus group 1 and student [4.1] students in general did not support the view that this difference in sound affected the performance and one student in focus group 2 felt the sound on VERT helped, "*I thought the sound was alright, actually at least you there was a link between what you were doing and what you were seeing. In that respect, yeah it was good.*" [2.1],

Another sense frequently mentioned was that of touch. Interestingly it was mostly not the ability of being able to touch the patient that was the issue. This was only mentioned by two students in two different focus groups who felt not being able to touch the patient made a big difference to the simulation on VERT, "*I wasn't all over the mannequins (Lots of laughter), but it made me feel a lot more detached, the fact that you couldn't get near to it even.*" [1.1] (*Lots of agreement.*) "*It would be better if you could touch the patient. You know, interact with it.*" [2.3]. Touch was recognised by other students in the focus groups as an issue for the VERT simulation, but not the issue of being able to touch the patient. Rather the main focus was on being able to touch and use equipment such as the couch controls and ruler which have already been mentioned, indicating some overlap of sub issues within these two main themes.

The final sense discussed by the students in relation to their perception of reality was that of sight. The general feeling of the focus groups was that they could see the setup better on the LINAC than they could on VERT.

[1,2] "*Yeah, I could see better at Bart's*."
(*Chorus of agreement*)
[1,2] "*Yes I found like I couldn't see where I was going really*."

This was also supported by student in focus group 2, with the main focus was on not being able to see the treatment distance (Focus to Skin Distance, abbreviated to FSD).

[2,1] "*FSD, Yeah. It was a bit more difficult to see in the VERT one.*"
[2,4] "*In the VERT one it was harder to see… because you couldn't read the numbers. I think I used the lasers*."

This difficulty with seeing the FSD was acknowledged by students in the other focus groups, student [3.2] stating, "*Getting the FSD right took a while, it was a problem on VERT.*" Student [3.3] adding, "*There was just a little bit of fuzziness of VERT*." In focus group 4 students commented that,

[4.2] "*You can't see the FSD. Is it just me or could you guys see the FSD numbers?*"
[4.5] "*I could see them but it was really hard.*"
[4.3] "*It was very difficult to see the FSD.*"

The issue with vision however went beyond not being able to see things such as the FSD clearly. Students also commented that the 3D visualisation in VERT was confusing. Students in focus group 4 commenting:

"*It's hard to know what kind of angle to look at it. You think that the 100 is there, but every time you move the distance seemed to change.*" [4.1]
"*No the movement did not feel realistic (Shakes head.) vision wise.*" [4.3]

This implies that not only was the distance hard to see, but also as the student moved the distance seemed to alter unrealistically, a view supported by students in focus group 1, "[*when you move in 3D I was so confused because the view would all change oddly*." [1.2], ([1.3/1.6] *verbally agree and nod*).

Of the five senses the only one not mentioned by the students in relation to the impact on their perceptions of reality was that of taste. Their experience of reality during the study was affected by the impact on each of the other four senses.

6.4 Theme #3 Learning opportunities

Despite the negative comments raised by students about VERT such as issues with the equipment and a certain lack of reality, they did comment favourably about the learning opportunities were also provided by VERT. The first of these was the opportunity afforded to the students to learn about how to use the pendant. With VERT there was an overwhelming agreement within the focus groups that the opportunity afforded them to learn about how to use the pendant was beneficial, "*I think it's brilliant for the pendant, pendant wise you get to use the actual pendant, which is the same as in the actual department so that's really useful.*" [3.2]. Again there was almost universal agreement within the focus groups that VERTs usefulness, particularly relating to using and the equipment especially the pendants is most beneficial if given before the students attend clinical practice.

"*I think the hand pendant's really the most useful thing, to see the hand pendant before you go into the department, so you have a bit of an idea about where the controls are.*" [3.2].

"*I think it's good for helping get you ready.*" [4.4]

As well as the pendant some students also felt that just being able to see how the machine moves benefitted them before going into clinical, "*It's good to see like, how the machine can rotate, like you might know now how the machine is able to rotate before you go into clinical.*" [1.4].

Although this view that VERT was useful as a learning opportunity prior to attending clinical some students did oppose this view,

[4.4] "*It gives you the feel of how you would use a real… like LINAC.*"

[4.1] "*I don't find it that useful.*"

[4.5] "*I didn't find it helped me, learning in clinical is best.*"

One student felt that the differences between VERT and real life in some instances helped you learn more effectively. When students were talking about not being able to use the ruler on VERT, the response from student [3.3] was "*that was OK, it just made you think a bit more.*"

Most students, in fact, wanted more simulation time, not only in preparation for clinical, but also with setups that are not seen frequently. The students generally felt that VERT offered an opportunity to practice the setup in preparation for seeing one in clinical such as electron setups. This opportunity to practice more using simulation students felt would positively affect their confidence, "*if we had the practice from VERT or even a LINAC then a person would be more confident in doing it.*" [2.2], but also their performance, "*it's like anything, the more time you have to practice surely the better you would get.*" [1.6]. Student [3.4] links these two issues and comments that the confidence gained from the simulation is linked to an increased performance in a real life situation, "*VERT is very much like, erm, it gives you more confidence actually being on VERT and when you are now live you have a measure of… you have a measure of confidence you never had before. And I think your performance is better.*"

When asked generally about the use of VERT some students considered that group work helped with their learning, "*I think working together helps.*" [3.1], "*Yes, we can help each other, look at what each other do.*" [3.3] Most focus groups felt that the group size could be smaller when VERT was used for teaching purposes as this gave a better learning experience.

Students also felt that having no time pressure would help as you could take more time which one student felt helped with the electron setups, "*when you are in that cloistered little room* (VERT) *it's, the stress is off, the pressure is off and I felt it was a lot easier to do it.*" [3.3] Another student commented that on the LINAC, "*everything's happening so fast I'm not taking that information in.*" [3.2]

Students also acknowledged that VERT may help with students who preferred a more visual learning style, "*people that are very visual and that would be and so when you just talk through a technique it doesn't you know, it's abstract, but when they can see it, it would really help those visual learners.*" [3.1] Some students also thought that VERT use should be extended as a demonstration tool to help show different techniques, not necessarily as an interactive learning tool, but as a way of the lecturer demonstrating the technique visually in lessons.

6.5 Theme #4 Assessment of competence

Despite many positive comments about simulation using VERT as a learning tool there was an overall feeling that of the two simulations only the LINAC simulation was suitable as an assessment tool. Student [2,4] stated, "*I think if it was a VERT competency I'd never get it signed off.*" Student [2,2] adding, "*I don't think that it's valid as a competency measure using VERT, erm maybe, maybe even for the model.*"

Students did not feel that their performance was comparable on both simulations, and most students who commented on this felt that their performance on VERT was worse than on the LINAC although some felt this was mediated by the sequence the simulations were completed in, "*When I'd done them both, it made me realise my VERT one was rubbish.*" [1.1] "*I feel the same, my Bart's one was a lot better than the VERT one.*"[1.2]. Many issues already identified in the preceding themes were cited as a reason for this, for example student [3.1] stated that, "*Being in the hospital helped, put you in …. Radiographer mode. You took it all more serious.*" [3.1]. This statement was supported by student [3.4] "*True… I think it helped me and I tried harder.*"

As suggested above reality also had a major role to play in why students felt the LINAC simulation could be used to assess competence when VERT could not. When the students in focus group 4 were asked, "*you also did an electron setup on a LINAC do you think you can use that as a measure of competency?*"

(All say yes, all nod heads. All say definitely.)
"*Yes you can use that it's really more realistic. You've got the machine there, the couch controls, you've got the hand pendant, you can measure, you can do a lot of stuff, but with the VERT you are just looking at the screen and just judging by luck.*" [4.5]

When asked if VERT could be used to measure competency students were clear in that that they felt it could not.

"*I don't think there's anything particularly wrong with it. It just can't be applied to real life.*" [2.2], "*It's just not the same.*" [2.4]

"*No, even then I don't, I don't think you can because all the virtual, virtualness isn't as reality. I think if you wanted to practice, use it as a practice method, very good, but for trying to sign off a real skin app for electrons then it isn't going to be very realistic and you're not going to get accurate competence.*" [4.3]

"I don't think that it's valid as a competency measure using VERT, erm maybe, maybe even for the model." [2.2]

"*It has to be a real LINAC*", [4.2 & 4.3]

More practice was also considered a major issue about using either simulation as a measure of competency. Students felt that being asked to complete a mock competency assessment when not ready was not fair. Student [3.3] commented that, "*You can't figure out why you can't achieve it!*" Other felt that it would only be a fair assessment if you could work more on VERT,

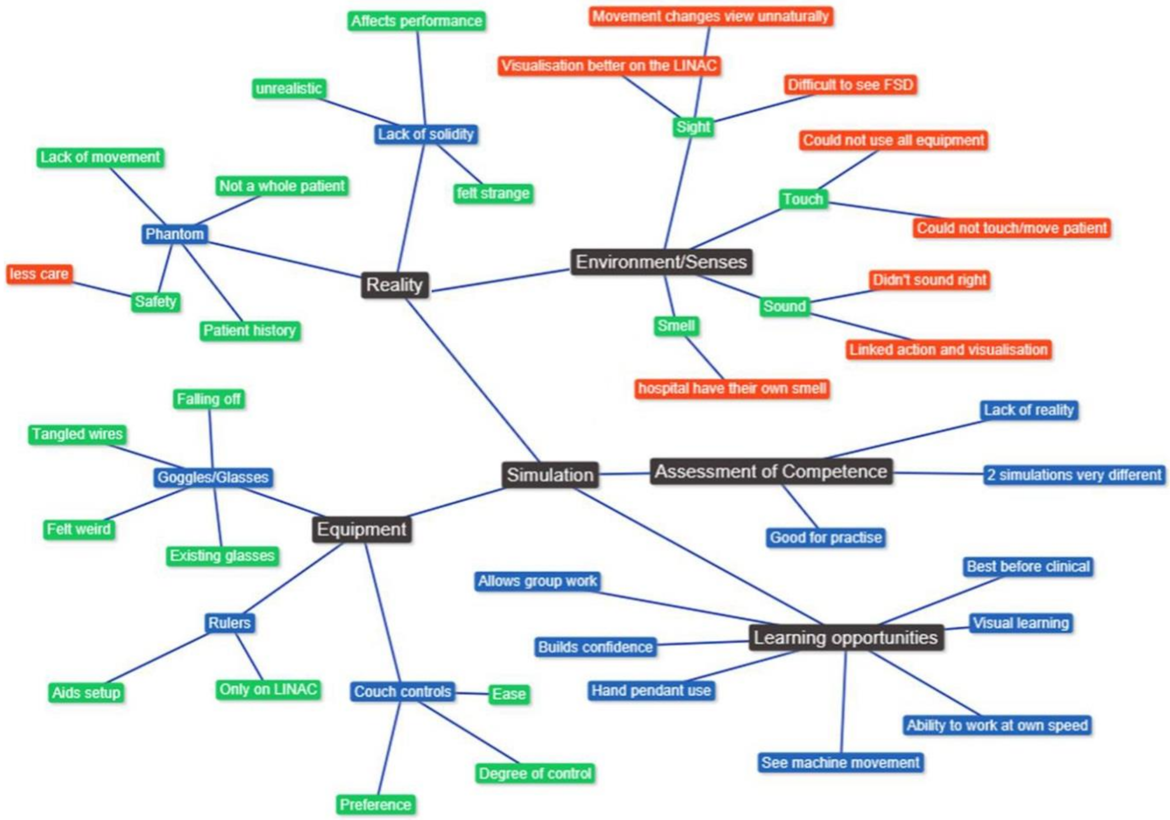
"*if I get used to how it works, then maybe, but, I dunno.*" [4.2],

"*If you get used to it and do a lot more skin apps."* [2.2].
"*It's probably not good as an equivalent, but in terms of practice. Yeah with practice.*" [2.4].

"*I suppose if you had more time to practice, it's like anything, the more time you have to practice surely the better you would get.*" [1.6]

Figure 6.1 below gives an overview of the themes presented within this chapter and how they relate to each other around the central concept of simulation.

Figure 6.1 Thematic analysis framework



This chapter looked at both the quantitative data and qualitative data generated by the research. This was attempted to be done in a factual manner, whilst pulling out the key points. The next section, the discussion, will pick up on the results presented here and discuss their meaning in the wider context of previous research findings.

Chapter 7 - Discussion of Results

7.1 Chapter Overview

This research set out to explore the use of virtual reality as a means of assessing clinical competence. The major outcomes of the study are presented and discussed in section 7.2. This section revisits the research questions individually in the following sections and puts it in context with other research in the area.

7.2 Discussion of the results for Aim 1

The first aim of this work was to identify if immersive tendency predicts the feeling of presence; the following null hypothesis was used, namely that there is no relationship between immersive tendency and presence.

As stated in the literature review, presence is felt to be a key factor in virtual reality and immersion is felt to be a key element of presence. Immersive tendency attempts to characterize the subject's susceptibility to presence experiences in advance of the experiment.

The results, however, indicated that a significant positive linear association existed only on the VERT simulation and then only for the ITQ subscale Game, the measure of the tendency to play video games. This was positively associated with the Total PQ score and the three PQ subscales; involvement, sensory fidelity and adaptation/immersion, Tables 5.5 and 5.6.

These findings differ from those of both Witmer & Singer (1998) and Ma & Kaber (2006) who found an identical positive linear association between the overall ITQ score and the subjective ratings of presence ($r = 0{\cdot}24$, $p = 0{\cdot}0001$), so indicating that individual tendencies as measured by the ITQ can predict virtual presence as measured by the PQ questionnaire. However, the findings are not consistent across all studies; and the consistency of the relationship between the ITQ and PQ scores has been questioned by other authors. For example Faas *et al*., (2014) found no correlation between immersion and presence, whereas Johns *et al*. (2000) in trying to replicate Witmer and Singer's findings initially found no significant relationship between ITQ and PQ scores for either high presence or low presence virtual environments although the correlations reported were higher than those reported by both Witmer and Singer and Ma and Kaber, $r = 0{\cdot}54$ and $r = 0{\cdot}62$. However, when the authors removed an outlier the correlation for the high presence environment became significant, $r = 0{\cdot}86$, $p < 0{\cdot}001$, so leading the authors to conclude that the relationship between immersive tendencies and presence as measured by the ITQ and PQ only hold true under high presence conditions. Part of the rationale for removal of the outlier was the fact that unlike others in the study this subject indicated on the ITQ that they were a frequent game player. The rationale for doing this is supported by the findings of Peña & Blackburn (2013) and Tortell *et al*., (undated) who found that users who play videogames frequently experience and interact with virtual environments differently than those who do not. This rationale is also supported by other authors such as Romano & Brna (2001) and Youngblut & Perrin (2002) who found a positive relationship between game playing experience and presence in virtual worlds, a finding that is also supported by the findings of this study.

The finding that there was no relationship between ITQ and PQ for the LINAC simulation may be explained by the concept mentioned in the literature review that immersion which Witmer and Singer propose is a key element in immersion in VR may not be generalizable to other media. Alternatively that presence is, at least in part, a product of the technology as suggested by authors such as (Bystrom, Barfield, & Hendrix, 1999; Draper, Kaber & Usher, 1998) and as the technology used in the two simulations is different the perception of presence will also be different in the LINAC compared to VERT, an issue is that is discussed in the following section.

7.3 Discussion of the results for Aim 2.

The second aim was to assess if presence would be higher on the LINAC simulation when compared to the VERT simulation. The null hypothesis being tested was that there would be no difference in Presence scores between the two simulation methods.

The tests shown in Table 5.7 indicated that the null hypothesis could be rejected for both the overall score and for all the subscales making up the instrument; the scores being significantly higher on the LINAC compared to VERT.

The largest difference between the VERT and the LINAC scores was for Sensory fidelity where the LINAC score was 54.6% higher compared to the VERT score. The Sensory fidelity sub-scale includes items on visual, auditory, and haptic items; all of which were highlighted in the focus groups as issues, particularly with the VERT system.

The questionnaire's authors (Witmer, Jerome & Singer, 2005) identified a moderate positive correlation between Sensory fidelity and Involvement postulating that simulations with a high Sensory fidelity are more likely to capture the user's attention, which in turn will support a greater level of user involvement. Poor sensory fidelity on the other hand would distract attention away from the task and activities focussing it instead on trying to process and understand the incoming information. Given that the study was looking at using the simulations; primarily as a means of assessment of competency, the intended users of the simulation should be experienced and should have acquired all the necessary skills that are needed for the assessment, see Figure 3.1. This would imply that for the simulation to be suitable for its needs a high fidelity needed to be evidenced, which was true for the LINAC, but not for VERT, the students scoring VERT considerably lower (23·8) compared to the LINAC (47·1) on the fidelity sub-scale of the PQ. Students also commented on the lack of fidelity of the VERT simulation when compared to the LINAC simulation in the focus groups. If we accept the earlier definition that fidelity includes different parameters such as environmental fidelity, task fidelity, physical fidelity, cosmetic fidelity, psychological fidelity and cognitive fidelity; then VERT fails on a number of issues, students commenting in the focus groups that the VERT environment "*didn't feel like a treatment room*". Students also mentioned that the sound and smell of the room being wrong and that the room didn't feel like a hospital room. There were also comments on the lack of reality as the VERT room lacked physical presence, the students being able to walk and look through objects, yet not being able to interact with some elements of the simulation such as the couch controls. They also had trouble when changing views, the movement not feeling right, which they felt did not conform to reality and they also had major issues with being able to see the FSD light clearly, an issue that did not arise on the LINAC. All these issues raised by subjects in the focus groups could explain the differences in the sensory PQ score.

The second largest difference identified between the two simulation methods was in the subscale "Involvement" which had a 34·9 percent difference, $p < 0{\cdot}05$. As the authors of the questionnaire indicated a positive correlation between Sensory fidelity and Involvement this finding is not unexpected.

It must be remembered however that the PQ questionnaire was designed and validated only on virtual environments. The use of the questionnaire has only been used to look at presence in a real life situation by one research group, Usoh *et al.*, (2000) who rationalised that the questions within the questionnaire were not unique to virtual environments. The authors looked at the Witmer and Singer, and Slater, Usoh, Steed presence questionnaires which were used to measure presence in two identical situations, one virtual, one real. The real life situation had a slightly higher score of 90·6 (S.D. 18·4) than did the virtual one 90·3 (S.D. 14·5) although the difference was not significant leading the authors to conclude that the questionnaire was unsuitable for cross environment comparisons. This finding is not supported by this study, which found a significant difference in presence scores between the two identical simulations, and also in all presence subscale scores, Table 5.7. One rationale for this might be that the feeling of presence is thought to be linked to task; Usoh *et al*'s study had the task of finding a red box in a laboratory, whereas the students in this study were commenting on presence related to an environment and equipment they would have almost daily exposure to. This familiarity with the environment and equipment may highlight any issues with the user's experience of presence in a virtual representation of the environment. As well as familiarity with the task it could be argued that whereas the task in this study was realistic, the task in Usoh's study of finding a red box in a research laboratory was not and this too might have affected the feeling of presence.

Immersion is part of the presence and within the concept of immersion is the concept of engaging the user's attention to the task, (Grimshaw, Charlton & Jagger, 2011). Attention involves sorting through information being received, selecting some for further processing and inhibiting other information from further processing, (Ashcraft & Radvansky, 2014). For more complex actions more concentration and attention is needed for their accurate execution, (Bedny, Karwowski & Bedny 2012) and the user is looking for more information from the system, being more keyed into the information being provided. In this heightened state of concentration users rely more on the information being provided by the virtual reality system and therefore are more aware of when this information is lacking, or not quite as it should be, so detracting from the feeling of immersion and presence. If true this would indicate that both the perceived realism of the environment and the perceived realism of the task and the task complexity are all important when considering presence.

In support of this hypothesis it is noted that the smallest difference in subscale score along with the Adaptation/Immersion score was Interface quality which had a 22% difference between the two simulation methods. The main method of interacting with the unit when performing the setup was through the handset which was identical in both the VERT and LINAC simulations. Students in the qualitative evaluation did not comment on the handset, but did comment on not being able to use the couch controls which were available in the LINAC simulation which might be the reason for the difference in score. The largest difference in score was seen in sensory fidelity, which the student picked up on in the focus groups as being a major issue in Theme 2, Reality. Within this theme subjects identified many issues such as lack of patient movement and only part of the patient being present VERT which failed to fully engage all the subject's senses, it did not "feel right" and as such the experience the students were subjected to moved further away from the reality of a patient's electron setup in the LINAC simulation.

7.4 Discussion of the results for Aim 3

The third aim was to try and assess which student characteristics moderate the presence scores, the null hypotheses that were tested being that neither age nor gender would affect the presence score.

Males had a significantly higher total presence score compared to females on both simulations. Males also had higher score on each of the four presence subscales, although interface quality was not significantly higher on the LINAC simulation and the sensory fidelity score on the VERT simulation was not significantly different, as shown in Table 5.8 and 5.9. If we consider the VERT findings in isolation the findings could be put into context by comparing the results with previous studies such as Felnhofer *et al*., (2012) who found that females had a significantly lower feeling of "Being There" when presenting to a virtual audience than males ($p = 0{\cdot}004$). There was also a difference in both Spatial Presence and realness observed with males reporting significantly higher values of Spatial Presence than females in both groups. It has already been identified in this chapter that there is a relationship between games playing and both immersion and presence and this finding might exist due to a difference in gaming experience and computer use. The 2005 Kaiser Family Foundation report indicated that boys between the ages of 8 and 18 were spending three times as much time as girls of the same age playing video games, (Roberts, Foehr & Rideout, 2005) and although this gap had lessened by the time of a more recent report, (Rideout, Foehr & Roberts, 2010), although boys were still spending twice as much time as girls playing video games. It is therefore not unreasonable when confronted with this evidence to suggest that the difference in presence scores between the genders is probably due to prior video game experience. One area that could be linked directly to previous games play is the case of the controller where males being more used to using controllers on gaming machines helped prepare them for using the hand pendant on VERT.

This line of reasoning except for controller use, however, should not apply to the LINAC simulation where males also had a higher presence score, unless you extrapolate the idea that simulation as a whole is being viewed as game playing regardless of the setting. There is no literature available to support any hypothesis relating to this aspect of the work as presence measures have only been used in real life in one previous study. However, this raises the possibility that what is perceived to be a gender difference in virtual presence may be more and extend into non VR simulations such as CPR training. This proposed gender difference of presence in real life simulation is supported by one study that noted a gender difference in attitude towards high fidelity simulation, males having a higher positive attitude scores toward the simulation than females, ($p < 0{\cdot}05$), which in turn may affect presence, (Grady *et al*. (2008).

No relationship was seen between the age of the student and presence within this study on either simulation, however some moderate and strong correlation values were observed on the LINAC simulation that approached significance. This finding is not unexpected as the study contained a relatively narrow age range of participants, between 19 and 42 years of age. The vast majority of students undertaking the undergraduate radiotherapy programme are school leavers, and over 65% of the subjects were between 19 and 24 years of age. Another reason why the result is not unexpected is that the evidence base to date for the possible effect of age on presence is very limited and inconclusive. Of the four previous studies that have reported on the relationship between age and presence, two studies, Van Schaik *et al.*, (2004) and Ling *et al*., (2013) found a negative relation between age and presence, $r = -0{\cdot}70$, $p < 0{\cdot}001$, and $r = -0{\cdot}24$, $p < 0{\cdot}05$, whereas Schuemie *et al*., (2005) found a positive correlation, $r = 0{\cdot}380$, $p = 0{\cdot}015$). The fourth study (Siriaraya & Ang, 2012) found that younger people had a significantly higher level of social presence than older people, $r = 0{\cdot}64$, $p < 0{\cdot}001$; and although younger people had a slightly higher level of physical presence (3·1) than older subjects (2·8) the difference found was not significant.

7.5 Discussion of the results for Aim 4

Arguably the main aim of the work, certainly at its inception was aim 4, which was to compare the simulation parameters of the two simulations to see if they were comparable and therefore whether VERT could be used as a measure of competency. The associated null hypothesis was that there would be no significant difference in simulation setup for the two simulation methods.

The first factor investigated was that of time. If students were undertaking the simulation in the same way it would be expected that the time needed to complete the procedure would be similar. Task time has been used before by authors such as Van Herzeele *et al*., (2007) and Khemani *et al*., (2012) as a measure of performance and also as a means to establish construct validity of VR simulation. Both the independent and paired analyses found no significant difference in time needed between the two simulators for the first setup which would seem to imply that both simulators are equally valid, however a difference was observed in the time taken for the second setup in the paired analyses and also for the overall time in the paired analysis, which would imply the opposite, that students performance was different on each of the simulators.

When looking specifically at the scores for the different setups noticeable differences occurred for all the setup components, students performing better on the LINAC simulation overall, see Figure 5.3, and in all aspects of the setup except dose variance in the second setup, Figure 5.2. The largest difference in performance between the two simulations can be seen in the FSD score for setup 1 where 96% of the subjects correctly set the FSD on the LINAC compared to only 47% on VERT. Students brought up the issue of FSD in the focus groups indicating that the VERT simulation had visual issues and were not able to see the FSD light well on the patient and also the problem with the distance appearing to change when you moved in the 3D world. Some students admitted to having to use the lasers because of difficulties in seeing the FSD light. Although using lasers to set the treatment distance to the patient can produce an accurate FSD to the patient if the required distance is 100cm it is not one that is frequently used in the clinical department the preference being to use the FSD light. The lasers are routinely used for aligning patients correctly on the treatment couch in terms of straightness and rotation, but are mounted so that they run through the isocentre which is positioned 100cm from the focal spot. In the case of electron setups as the required distance is 100cm FSD the method is valid, but not every student thought to use this method, which probably accounts for the low average for the FSD scores in setup 1 for the VERT simulation. This adaption of technique on VERT is an example of affordance, the students taking a tool designed for another function and putting it to a use to make up for an inherent inadequacy of the VERT system.

The degree of difference observed in the second electron setup for FSD was much reduced. Although not recorded by the researcher it was anecdotally noted that more students used lasers to aid in the setting up of the distance for the second setup. Another factor that may have contributed to this finding is that of reality, identified as theme two of the qualitative data. In general students thought that the simulation could be improved by having a more complete phantom so that it represents the whole patient. In this case the difference in performance between the two setups may have been related to the position of the setups; see Appendix IX for an image showing the approximate position of each electron field. Setup 1 was positioned on the anterior chest wall away from the patient's head, whereas setup 2 was located in the supra-clavicular area, far closer to where the head of the patient would have been if present. Therefore in a real life situation for setup 2 the radiographer would have to consider the position of the patient's head in order to try and avoid hitting the area. Looking specifically at the student's comments that might relate to this issue one student in focus group 2 commented, "*Yeah, in terms of setup it is similar, although the Bart's one felt harder because of the chin getting in the way*." This implies that even though not present the student is considering the chin placement on the LINAC, but as VERT was not mentioned this was probably not the case on VERT despite the simulated patient being identical to the one used on the LINAC. This is supported by a second quotation from a student in focus group 4 who stated, "*When on VERT, when I was doing the matching, if there was a patient's head there I probably would have taken it off.*" This statement is almost the reverse of the previous statement, the student admitting to not considering the patient's head position on the VERT simulation when doing the setup because if they acknowledge that they would have had a collision between the applicator and the patient.

This line of reasoning about thinking of the patient as being real could also be extended to consider dose variance. Dose variance depends on the difference in the standoff in each of the four corners of the field. Students as part of the setup would attempt to gauge the distance and then change the machine parameters such as gantry angle and couch rotation in order to try and reduce the variation. This action depends on very good 3D rotational perception which overall was considered lacking by students in the VERT simulation, yet for the second electron setup VERT scored higher than the LINAC. The students here appear to be considering the applicator placement in relation to the patient's anatomy more carefully and realistically on the LINAC so moderating the final setup parameters compared with the final setup on the VERT simulation.

Another way of looking at the results from this part of the study is to consider it as a test of concurrent validity. The Bland Altman plot shows that only three subjects managed to replicate their total score and have agreement on their performance measures on both simulations. The performance on the LINAC simulation was better than that on VERT, with subjects scoring on average 2·6 higher on the LINAC simulation compared to the VERT simulation and only three subjects scored higher on the VERT simulation than they did on the LINAC. The magnitude of this difference (2·6) is important given that the scale only ran from 0-12 and therefore this value represents a difference of approximately 22% of the possible score. However, it must be remembered that the LINAC simulation is not the gold standard. It is not how they would perform in real life, but it is a long established training method which is much closer to reality than the VERT simulation, a fact that is supported by both the presence scores and student comments. With the differences observed in the outcome measures investigated, time to undertake the procedures, overall score of the two simulations and the sub-scores making up this measure the conclusion must be that there is no construct validity between the two simulations.

7.6 Discussion of the results for Aim 5

$H0_5$ Cohort, gender and immersive tendency will not affect the simulation score.

Although not significant the cohort scores for the LINAC simulation increased for each year as they progressed through the programme whereas the VERT scores did not, (Table 5.20). The validity of simulators has often been undertaken by comparing groups with different experience, such as novices and experts, (Gavazzi *et al*., 2011; Schreuder *et al*., 2014). The premise is that the groups consisting of more experienced users should obtain a higher score than a group of inexperienced users. In this study each successive cohort has more experience with electron setups than the previous one, so giving us a novice, intermediary and expert group. Although perhaps the claim of expert for the third year cohort is a bit strong they certainly should have more expertise that the second year group that forms the intermediary group. With the LINAC simulation the more experienced the cohort the higher the mean score for the group. There were relatively small amounts of data especially in the least experienced group which only contained 3 data points. These low numbers could explain why significance may not have been achieved as the power of the test would be very low. Although the same argument for lack of significance could be made for the VERT simulation scores, what is noticeable is that the pattern of the cohort's scores are not as we would expect, the most inexperienced group having the highest score. The results therefore seem to indicate that construct validity may exist for the LINAC simulation, but not for the VERT simulation. This finding is worth reinvestigating as this would seem to suggest that content validity might exist for the LINAC simulation if the element of the study were to be re-attempted with larger and more balanced cohort sizes, so giving the test more power.

Various authors have reported that there is a gender difference in 3D spatial ability and spatial skills, (Marunić & Glažar, 2014; Ventura *et al*., 2013). It is now generally accepted when looking at spatial cognition that gender differences exist, especially in the area of mentally rotating 3D objects and spatial navigation where males outperform females, (Voyer, Voyer & Brydon, 1995), although it has also been noted that females show an advantage in spatial abilities relating to memory for object locations, (Levin, Mohamed & Platek, 2005). The differences appear to be derived from a mixture of biological factors such as genetic predisposition and environmental reasons, (Miller & Halpern, 2014). One common cited environmental factor for this difference is that the early recreational activities tend to be different for each gender, (Feng, Spence & Pratt, 2007).

Organised movement is impossible without an individual first acquiring, storing, and manipulating spatial information and because of the difference in spatial cognition between genders we may expect to see a difference between genders both in terms of time taken to undertake the setups and the score, however none were observed. Although there may be a difference in spatial cognition in adults of different genders it does appear that training can improve spatial cognition in both sexes and that the learning trajectory of females is comparable to that of men and gender parity in 3D spatial performance is achievable after training, (Spence *et al.*, 2009). The use of virtual environments specifically has been linked to improved spatial cognition. Meijer, Geudeke & van den Broek (2009) showed that VR could be used to increase spatial knowledge and that this was increased in a photorealistic environment compared to a non-realistic environment, the fundamental properties of VR, interactivity and three-dimensionality being crucial to spatial learning, (Stanton, Wilson & Foreman, 1996).

Overall males were seen to be quicker than females by nearly three minutes on the LINAC and 30 seconds on the VERT although neither of these differences was significant, Table 5.18. The lack of significance for the LINAC setup time of 2.84 minutes is probably due to the large natural variation in time taken observed during the study, the setup times ranging from 3·3 minutes to 24·50, standard deviation 4·6 minutes. With overall setup scores males marginally outperformed females on both the VERT and LINAC simulation although there was no clear pattern within the two different electron setups. Males outperformed females on the first electron setup on the LINAC, but not the VERT, the gender performance being reversed in the second electron setup and in the LINAC electron setup female's significantly outperformed males. This would suggest that this study cannot support the commonly held gender differences discussed above, although the number of males in the study was low so affecting the power of the text and a more balanced study design might be needed to further explore this relationship. Two other points also need to be considered in relationship to gender differences. Firstly the imbalance in the study design in that most students, approximately 86% of those undertaking the setups, were year 2 and 3 students, both of which had had considerable clinical time and simulation time working on similar spatial tasks and it may be that any difference that may have occurred due to gender differences in spatial cognition at the start of the programme may have been removed through training. Secondly, although not asked about directly, the issue of gender was not raised during the focus groups by the students and therefore the importance of gender as an issue was not one that was identified by the subjects.

7.7 Discussion of the results for Aim 6

Aim 6 looked at the way the units were being used by the students and asked the question, do participants utilise the same cognitive process on both systems?

Looking first at setup 1 subjects on both simulations predominantly used the longitudinal and vertical couch movements in the first 30 seconds. This is understandable as the starting point for both simulations was identical with the couch low down near the floor and moved out from the gantry as it would be in real life so as to allow the patient to easily get onto the couch and safely lay down without risk of injury from the unit. All the early movements are related to getting the patient into the correct treatment position and it is easier to see which movements are needed when the applicator is near the patient and so other movements such as the collimator rotation and couch rotation would naturally follow later. What is noticeable is that many of the movements appear to start earlier on the LINAC, and the use of the couch vertical and longitudinal movements after 30 seconds are starting to decrease. This could suggest that students on the LINAC are more quickly seeing what needs to be done, and then completing the process faster than on VERT although Figure 5.10 does not support this as there were more periods of inactivity on the LINAC than on VERT. A more likely explanation would be that students could not only use the handset on the LINAC, but also the couch controls and making the use of both the couch and gantry movements at the same time which allowed more movements to be made at the same time, which is supported by the data in Figure 5.12 which shows that on the LINAC students were far more likely to use two movements at once on both the first and second electron setup, and that the most commonly paired movements were couch vertical and longitudinal. Also on the couch controls not only can the couch be moved using the motors so replicating the pendant control it also has the ability to “free” the couch top, basically take off all the locks on the longitudinal and lateral couch movements. This allows very rapid coarse movements of the couch top so allowing a speedier “rough” positioning of the patient under the electron applicator. This might also account for some of the variation observed in setup times in section 5.7. This idea would be supported by information from the focus groups where students raised concerns about not being able to interact with the couch controls, which is discussed further in section 7.81.

In the second setup the starting point was the end point of the first setup and as such was not from a common point in space as it was in setup 1. It is not uncommon for some patients to require two electron fields and the second electron treatment represented a second field on the same patient. Because the start position for setup 2 was slightly different for each student this will lead to some variation in the use of machine moves that were not seen in the first setup giving greater variation in the movements observed. Another variation seen here is that some students tried to complete the second setup with the floor rotated the same way whilst others decided to move the couch floor rotation through approximately 180 degrees for the second due to the treatment field being on the other side of the patient whilst others did not. Again there was similarity in the main movement at the start of the second setup and the fact that it seemed to start ending sooner on the LINAC than on VERT. In this case the main movement was a couch vertical movement, students moving the couch down in order to more safely move the patient into the correct position, reducing the possibility of the patient being hit by the electron applicator of the LINAC whilst moving to the next setup. Both the main movements at the start of both setups would be considered as expected movements for most setups of this type. The movements being undertaken are not fine movements to get the patient into the exact position and so don’t need good visualisation, but coarse movements that either get the patient roughly in position or allow other movements to safely occur. It is interesting to note at the start of the second setup students on both the VERT and LINAC simulation dropped the couch to protect the patient. This seems not to coincide with the argument put forward earlier when considering aim 4 when it was hypothesised that the difference in setup 2 could be explained by lack of concern for patient for patient safety due to lack of realism, however this did not consider procedural memory, a set of action rules used to perform familiar or routine tasks and generally resides below the level of conscious awareness, (Wang *et al*., 2014). The movements made at the start of both setups are standard, particularly those at the start of setup 1 where this combination of movements would be used for virtually every patient treated and so could be being done subconsciously, whereas the consideration of later moves in the electron setup tend to be more unique, and need greater consideration, depending to a large extent on the area being treated.

When comparing the movements in the final 30 seconds of the setup the movements are more confusing and it is harder to discern a clear pattern between the two simulations. The movements here are finer movements that are being made to try and achieve the correct setup and are more dependent on being able to see the marks on the patient and beam parameters.

What is noticeable is the greater use of the couch vertical movement on the first VERT setup at every time point in the last 30 seconds compared to the LINAC setup. This is probably due to the difficulty in observing the distance light identified by the focus group data, as this is the main movement responsible for setting the FSD. The difficulty of seeing the FSD might further affect individuals in terms of the stress or anxiety they are experiencing. Students will be stressed to some extent in both simulations, the stress arising from both the task and the observation. This stress experienced by each individual should be the same on both simulations if the task and simulation are the same, but it has already been established that the students experienced more issues with the verity of the VERT simulation, particularly in respect the FSD than they did on the LINAC simulation. Although the students were not specifically asked in the study about stress, some students did mention it in respect to lack of time pressure, but not specifically in relation to their performance, although it may have been experienced at a subconscious level. According to Fielding, Shapiro & Tang (2012) stress has been shown to enhance an individual's selective attention to a stressor, so impairing their performance in tasks requiring divided attention so when it came to finish the setup and get the distance correct this factor might have come to the fore and subjects ability to consider other factors such as rotation and standoff, which also scored lower on the VERT simulation may have been reduced. Stress and anxiety also affects cognitive reasoning, the decision making strategies being employed being considered more suboptimal, (LeBlanc, 2009) and if students were experiencing stress from an inability to perceive the FSD or move the bed as they wanted this might have affected their actions during this part of the simulation. This supposition would fit with Harvey *et al.'s*., (2012) findings that some aspects of simulator performance appear to be compromised in complex clinical scenarios due to the subject's demonstrating elevated subjective and physiologic stress responses. Further evidence in support of this supposition comes from Jaeger & Adair (2012) who found a negative correlation between ease of use of the simulator and perceived stress.

Another difference to highlight is the difference in use of gantry and couch rotation movements, particularly at the end of the second setup where on the LINAC the use of both are roughly equal whereas on VERT there was a much greater use of the gantry rotation. The use of couch rotation and more especially gantry rotation movement is predominantly associated with trying to ensure equal standoff in each of the four corners of the treatment field. The implication therefore is that on the LINAC simulation more subjects are satisfied with their standoff earlier than on VERT, where consideration of standoff is dominating a lot of subject's thoughts and adjustments are still being made.

Although noted earlier that having couch controls as well as a hand pendant allowed multiple controls to occur at once, and that most multiple movements did involve at least one couch movement, not all multiple moves were couch related. There was a general propensity to use joint movements by students when on the LINAC compared to when on VERT. This might be linked to either the verity of the VERT simulation being not as good as indicated in the focus group interviews. If the subjects were not seeing the image as clearly and the movement did not look right they might take the process slower ensuring that the one movement they were doing was right before moving onto the next.

The other factor looked at in this section was machine inactivity, which tended to be higher on the VERT setups. These periods of machine inactivity probably occurred as the subject was assessing the setup and considering the next movement. This difference was not seen to be that great in the first 30 seconds or final 30 seconds of the setups so must have occurred in the middle part of the setup as large differences in inactivity were seen between the two simulations, Figure 5.9. There did appear to be some relationship between inactivity and the time taken for individuals to undertake the assessment, with moderate to strong positive correlations in three of the four setups undertaken (Table 5.21) which would appear to be a logical finding. A number of possible explanations could be put forward to explain why there were more periods of inactivity on VERT compared to the LINAC. Firstly, because of the difference in fidelity between the two simulations it could be postulated that more thinking time was needed on VERT to assess the current situation. Because of the lack of fidelity the subjects were not being able to see the setup parameters and patient as clearly as they would like due to issues in visualisation and feedback such as not being able to touch the patient and use rulers which are areas highlighted by students in the focus groups.

Also feeding into this difference might be the issues with the glasses, some subjects commented on getting tangled in the wires and more commonly that they kept falling off. It would be logical inference that if this did happen, then in order to retrieve and reposition the glasses the student would have to stop using the controller until repositioned. This issue is extrinsic to the simulation, but intrinsic to the simulator and would be an issue for any simulation done on VERT in 3D.

Less easily explained when looking at the data presented in this section is the fact that most student inactivity was seen on VERT and students on VERT were also less likely to use multiple movements of the machine parameters, yet it was reported when looking at Aim 4 that the VERT setup was on average performed considerably faster than the LINAC setup.

7.8 Additional Discussion of Qualitative Findings

The data gathered from the focus groups provided further details to the data gathered from the simulation assessment. Additional to providing more detail and explanation to the quantitative data which has already been looked at the analysis of the qualitative data raised four broad themes. The data will be looked at in order to look at the final aim of the project, what do students think about the appropriateness of both methods as an assessment tool?

7.8.1 Equipment use

The transcripts from the four focus groups made it clear that the students frequently had problems with the equipment; in particular they had major issues with the active shutter 3D glasses. The use was hampered by the students' need to frequently alter their head position to get a good view of the setup, which led to two main issues. Firstly keeping the glasses in place during the setups and secondly getting tangled up in the lead from the glasses to the positional receiver which was usually attached to the students' belt or pocket. Constantly needing to adjust and keep control of the 3D glasses could have detracted from the experience of VERT and phrases such as "can we change the glasses?" and "working without the glasses was best." would seem to support this. The glasses provided by the manufacturers were quite bulky and heavy and recently different manufacturers have produced lighter models than more closely resemble normal glasses which are lighter and easer to wear and it would be easy to swap the old glasses out. Another possibility would be to replace the position tracking monitor off the glasses which adds to the weight of the glasses and use some form of headband to hold it in place.

Another problem with the 3D glasses was found by students who usually wear glasses, who found it uncomfortable to wear the 3D glasses over their own glasses. Again lighter glasses suggested above would afford some benefit to students who experience this issue.

A second issue with the equipment was that of the controls. As stated earlier in the discussion the students had the choice to use the couch controls in the LINAC room, but not in VERT. which might have accounted for some of the difference in the timings between the students on VERT compared to the LINAC, one student commenting that having the couch controls on the LINAC simulation "*makes a big difference*". This may also have impacted on the presence scores as there may have been a preference for some students to use the couch as opposed to the pendant as indicated by the student above, which in turn might have led to frustration on the part of the student and heightening lack of reality and interfering with immersion. There is the option offered by the manufacturers of VERT to use an iPad with an app that simulates the couch controls. The students were aware of this app having used it in previous teaching sessions, and at least one student commenting on it in the focus groups. However, it was not used in this study as, as stated earlier it was felt that it would interfere with the setup as it had to be held rather than being fixed to the couch as in real life, which would have meant "juggling" two controllers during the setups. The decision not to use the iPad controller would have had only a minor impact on the timings, as although the app provides a photo-realistic iPad controller, it only allows you to use the motorised couch controls, which operate in much the same way and at the same speed as the gantry controller. The button that would allow quicker free movement by unlocking the couch top as on a real LINAC does not function in the app as the operator cannot physically interact with the couch in the VE in order to allow the movement.

The final issue raised by students relating to equipment was the ability to measure the amount of standoff. The VERT system being virtual did not allow the user to measure the standoff in each of the corners whereas on the LINAC students could put their rulers next to the applicator and see the variance of standoff in each of the four corners. This was not done by all students, but was by others and should improve the dose variance seen across the field.

7.8.2 Reality

Throughout the focus group discussions students frequently made reference to the reality or verity of the simulations. The first point raised in the results section applied to both the VERT and LINAC simulation and related to the phantom, which neither moved nor represented a whole patient. Movement of the phantom cannot be addressed easily in either scenario. In VERT although the machine and couch moves it is not possible to move the patient from with the simulation. Minor movements are possible with the software, but these relate to small longitudinal and rotational movements of the patient which are not the movements required, as one student stated "*I would just move the patient's chin*". This would be a simple movement in real life achieved by asking the patient to move their head to the left or right or to tilt their head up/down. This issue was particularly relevant to the second electron setup which was close to the neck area of the patient. Although not possible in VERT this might be possible in the LINAC as jointed mannequins could be used rather than a phantom which would allow lifelike movement of the simulated patient. Students felt that using a more life-like patient would raise their awareness of the issues related to the setup.

Whereas issues about machine movements cannot be easily addressed on both simulations equally; using a whole patient can be with little difficulty. In the case of a LINAC based simulation it would mean transferring a whole body phantom from the university to the hospital when needed. This would cause minor problems with transportation as anthropomorphic phantoms being made of tissue equivalent material weigh roughly the same as a person, although lighter models exist such as PIXY® which weighs approximately 48kg. In VERT the software is delivered with a whole body data set including electron fields. This wasn’t used in the study as the data is from the visible human project and as such there was no exact phantom match for the LINAC simulation. The student’s desire for a whole patient to be represented is at odds with the findings of Owen *et al*., (2006), who found no significant advantage in the use of whole body mannequins, although the study only compared performance and did not consider the subjects’ views. However, another more recent study that would lend weight to this argument found that subjects performed better on medium fidelity (partial task trainer covered by life sized poster of a pregnant woman) and high fidelity simulators (students acting as patients with the partial task-trainer appropriately positioned between their legs) compared to a low fidelity simulator, (just the partial task trainer). The study demonstrated better learning outcomes in the medium and high fidelity simulations compared to the low fidelity one, the only difference between the low fidelity and medium fidelity simulation being a cloth poster of a patent being draped over the simulator, (Brady, Bogossiana & Gibbons, 2015). This finding highlights that the psychological issues of simulation can be as important to the users as much as the simulator itself.

One focus group also suggested the use of “mixed reality” to improve the realism of the simulation, combining elements using real people with elements of virtual simulation. As with the patient they viewed the simulation holistically, wanting to include all elements of the patient’s treatment, not just the setup. Realism for the setup would be increased if it was linked to identifying the patient and taking the patient into the room. A realistic introduction to the setup would convince the student to take the assessment more seriously.

The need for a more complete scenario might be linked to the concept of contextualized learning, the mind naturally seeking meaning in context by searching for relationships that make sense and appear useful. Contextualized assessment, which has one dominant theme in that it should reflect “real-life” tasks and require students to utilize higher order thinking skills, (Klassen, 2006). Linked with this concept is consideration of part-task simulation with whole task simulation. Part task simulations are widely used for physical examinations such as rectal examinations and in life saving scenarios such as CPR, but they are focussed on only part of the whole task, at no point can a mannequin teach social skills, (Siebeck *et al*., 2011). Whole task simulators are supposed to offer a more authentic experience as they consider all components of the task. Kolozsvari *et al*., (2011) suggest that part-task training gives most benefit when the whole task is complex, and the task components are not highly integrated whereas whole task training appears more suited to when the task components are highly integrated and it would appear that at least some students are seeing the task holistically wanting both to interact with the patient and complete the technical side of the setup together integrating the two tasks together and if the simulation can be adopted bringing in more aspects of a realistic setup, making it more holistic, the students might engage more and their performance might be truer to the real situation.

A number of students and focus groups commented on VERT stating that, "it felt different…" and in some instances smell, touch and sound were raised as the reason why. When working on a task that is repeating the order of events from a previous experience the part of the brain engaged is the hippocampus that contains a number of cells known as place cells and time cells, (Eichenbaum, 2013). These place cells are involved in storing and/or retrieving spatial memories, whereas time cells encode moments in temporally structured experiences, both are essential to spatial navigation. A recent study observed that when rats moved in a virtual world their sense of space was not normal in that the firing of their "place cells" was different compared to when they were in an identical real space. They concluded that visual stimulation is not enough to give a true perception of the body's position in the virtual environment in relation to the surrounding objects. They suggest that other sensory cues needed to be present, and/or that incorrect sensory cues conflicted with the visual cues in VR as these also influence place cells. One potential consequence of conflicting information from the different senses was loss of spatial selectivity in VR, (Ravassard *et al.*, 2013). This might explain in part why presence scores and competency scores were higher on the LINAC compared to VERT. On the LINAC setup not only were the visual stimuli correct, but other subconscious stimuli being received by their other senses were telling them they were in a treatment room and the place and time cells would be producing the correct spatial memories helping them with the setup whereas on VERT there was a slight discord with what they could see and what their other senses were telling them so possibly affecting their retrieval of the correct spatial memories.

7.8.3 Learning Opportunities

VERT was considered by many students in the study to be beneficial as a learning tool, giving them more confidence by preparing them for clinical which they felt would help them perform and achieve better results. Simulation training has been shown in other research to improve confidence, (Curran *et al*., 2010), however, the improvement in Curran's study was no greater than that seen by the researchers control method of a training video.

This students' stance in this study reflects that commonly held by many authors, (Lateef, (2010); Maagaard *et al*., (2011); Aggarwal *et al*., (2009)) that simulation, which can include VR is a technique for both practice and learning. The students' belief in the usefulness of the simulation as a learning tool for electron setups is also important to its use as Huang, Liaw & Lai, (2013) showed that perceived usefulness was a positive predictor ($p < 0{\cdot}001$) of the learner's behavioural intention to use VR learning. This link between learner attitudes towards 3D VR systems and their use is also supported by other authors, (Sun & Cheng, 2009; Verhagen *et al*., 2012).

Another possible benefit of VERT identified by students was the lack of stress and time pressures when using the VERT system for learning. As discussed earlier in this section students had various issues with the VERT simulation and it was proposed then that this may have put stress on the students, which in turn may have affected their performance when undertaking the assessment. Here the students are acknowledging that when learning on VERT they are less pressurised than when working in the department, primarily because of the time constraints that exist in clinical. This supports other research such as Mathew (2014) who states that allied health professionals perceive time constraints as a significant barrier to simulated learning in practice.

The ability to practice on VERT with a lack of time pressure is seen as a positive pressure that might allow students to learn at their own pace; unaffected by the clinical pressures of needing to treat patients on time and perhaps without the emotional involvement people have in the department. The influence of stress and time pressure on performance is debated. For example, Goodie & Crooks (2004) in a simple probability matching task, showed an increase in both heuristic processing and performance if the subjects were asked to perform the task under a time pressure. This view is supported by Westlake, (2011: 4), who states that, “time pressure helps interactions and results.” Opposed to this view are authors who view time pressure negatively. One study looking at critical care simulation reported that time pressure significantly lowered the ability of the users; the user’s being less likely to intervene and missed more missed responses, (Thompson *et al*., 2008).

Regardless of the time pressure, students felt that the VERT simulation would give them more confidence and most focus groups showed a desire for more simulation time on VERT practising for clinical. This is important in experiential learning as many authors suggest that motivation is a key driver in experiential learning, (Saenz & Cano, 2009). One student [3.3] felt that the differences in VERT to real life helped him think a bit more about what he/she was doing. This could be beneficial to the learning process, which is perhaps why students felt that VERT could be used for teaching, but not assessment.

Interestingly many of the students and all of the focus groups commented on the use of the pendant, and it was this that was seen as being most positive about the simulation; learning how to use the pendant. The pendant was the only part of the VERT simulation that was real and identical to that on the LINAC and perhaps reflects that interaction is a key element to the subjects. When interacting with the pendant subjects are happy to observe the movements occurring on the machine as this is very similar to real life, but lack of interaction with the couch, patient and room become issues for the student as they limit interaction with the virtual world compared to the real world or the LINAC simulation.

7.8.4 Assessment of Competence

It was noted that student’s generally had a negative attitude towards the use of VERT compared to the LINAC as a tool for assessment. The issue of perceived usefulness has already been raised in a positive way when looking at VERT as a learning tool, however, here the connotation is negative. As well as being a predictor of intention and attitude towards the simulation according to Fetscherin & Lattemann (2008) perceived usefulness also plays a significant role in explaining a user’s acceptance and behaviour in virtual worlds. As such their lack of perceived usefulness of the VERT simulation in comparison to the LINAC simulation for assessment may explain some of the difference in the outcome of the qualitative study.

The exact reason as to why the students felt the VERT system could not be used was quite vague, although it seemed to revolve around realism. Whereas for learning reality was not seen as so much of a problem, reality was considered essential for assessment and any deviation from reality detracted it in their eyes as an assessment tool. This difference in student perception for the two uses of VERT could probably be related back to the theories of teaching where the differences in reality are not so important, the situation being active and in context, the differences as one student put it making them think more about what they were doing so encouraging reflection which might not happen if done on a LINAC as it would be more automatic, i.e. the students were less immersed in the simulation and had to think more about what they were attempting. This also might reflect why there were more time periods where there was no activity on VERT compared to the LINAC, the students stopping and assessing the situation trying to work out their next move. For learning, these small differences might actually help the student reflect on what they are doing and working out how they intend to go about it.

It is clear from the comments with the separation of learning from assessment that the students only considered assessment as being summative separating it from learning. No student considered learning through the assessment of competency.

The conceptual clinical competency pyramids have "showing" and "does" on the top two layers. As Carr (2004) pointed out the ability to demonstrate that they can do is not the same as does when faced with a real life situation. And as stated earlier both these methods are asking students to "show" and so this may not relate directly to "does". In any assessment it is important that the user and designer believe that the assessment is valid and a fair test. In this study it would appear that the students felt that the LINAC was better at achieving this aim than VERT.

7.8.5 Presence, Immersion and Competency

Immersion and Presence have been discussed earlier in this chapter, specifically looking at the results in light of the original aims of the study. The aim of this section is to pull the earlier discussions together and look at the concepts of both immersion and presence more generally and also in relation to the competence scores.

Much is made in the literature about the link between immersion and presence. As stated in sections 3.5.2 and 3.5.3 different authors have different opinions about what exactly constitutes these constructs. Regarding immersion, the generally accepted stance is that immersion is defined in technical terms, being a product of the technology that is being used. Immersion is a key feature of presence, the belief the user has that they are in the virtual world rather than the real world, (p: 40). The expectation, therefore, might be that the more immersion is present, the more presence is felt by the user and the closer the user will behave and perform as they would in the real world.

If we accept this relationship between immersion and presence we can see why students felt lower levels of presence in the VERT than in the LINAC simulation. Firstly the fidelity of the VERT simulation was not considered to be as good as in the LINAC simulation and secondly the senses in VERT were not fully engaged by the system. This type of competency requires significant visual engagement. Vision is the main sensory input utilised in this type of activity and the VERT system failed on a number of counts to actively engage the user. Referring back to affordance theory (Section 3.6.1) there were also a number of false affordances offered by VERT. These were mainly linked to the couch and although all the buttons present on the LINAC were replicated in the VE none functioned and so elements of functional affordance were missing. Hartson (2003), also talked about hidden affordances which were again present in the VERT system where an option exists that allows the user to look at the distances from each corner of the electron field to the skin's surface which might have helped the student ensure a more equal standoff. However, this option is hidden in a sub-menu and so was not used by the students. Also the distances do not automatically update as you move the equipment and so each time a distance needs to be checked the tool has to be switched off then back on again. Both the false and hidden affordances did have the ability to affect the user's performance to a lesser or greater degree and the false affordances were mentioned by the students in the focus groups as a difference, therefore these aspects of affordance theory should be considered when designing virtual environments as they impact on the feeling of presence.

Although vision was perhaps the most important sense utilised in the setup other senses were also engaged in the LINAC simulation, but not wholly in the VERT simulation. These included touch, smells and sound which the VERT simulation could not adequately recreate. Because these sensory elements were not delivered by the VERT system in relation to their equivalent real-world sensory modalities, VERT in the student's opinion lacked immersion, which in turn affected the student's ability to suspend their belief that they were working in a teaching room watching and interacting with a computer generated image. Immersion was not directly measured in the study, but rather the information relating to immersion came from the qualitative part of the study in the form of the subject's discussions in the focus groups about the simulations, which to a large part revolved around the equipment and what VERT could offer in terms of the interaction between user and system. Most of the discussion points raised in Chapter 6 relate directly to the equipment and hence immersion or how they felt during the simulation as a result of this difference, presence. Equally the students felt that many of the points raised affected their performance, particularly in VERT and this in turn might have affected their competency score. As well as commenting directly on the equipment they also highlighted that not only the senses that directly impinge on the task were important, in this case vision and touch; but other senses also need to be engaged in the simulation for the user to suspend reality and believe in the fictional world being created by the system in order to become immersed. Therefore other senses beyond those needed to directly interact with the system in order to undertake the task may also impinge on the user's ability to perform a task as they would do in real life. This engagement of senses other than sight is an often left out consideration of immersion and presence. Bowman and McMahan (2007) for example consider only visual immersion in their paper and look at immersion only in terms of visual factors such as frame rate, field of view and rendering, yet the subjects in this study needed a more complete sensory support in order to suspend their belief. Zhang & Fu (2015) described an experiment where the use of background music during an immersive task increase the participants' level of immersion and the sound and smells of the department would appear to have a similar effect according to the students.

Correlations were performed between the ITQ and competency scores, none of which reached significance. However, it must be remembered that the ITQ questionnaire was looking at the tendency for a subject to become immersed, not the immersion experienced on the system and so this lack of relationship could not be used to refute the earlier suggestion from the qualitative aspect of the study that links immersion to competence. Furthermore it must be noted that immersive tendencies was one of the variables used when randomising patients into each arm of the study, therefore the immersive tendencies of the subjects in each arm of the trial were overall very similar, (see page section 5.2.1), yet despite this the difference in presence between the simulations was very large (39·6).

The correlation between a student's competency score and their presence score was very poor with only one measure reaching significance on the LINAC simulation and the correlation values all being classed at best as "weak" ($R < 0{\cdot}3$) for both simulations. Although it can't be said that performance is totally independent of presence the evidence of any relationship from the quantitative data is very weak.

If we again consider what presence is; that it is a level of consciousness, the psychological sense of, "being there" it is logical to assume that if one believes oneself to be somewhere and something happens one will respond and act as one would in real life, which is how Zahorik & Jenison (1998) describe the concept of presence. Presence is therefore considered to be more closely related to behaviour rather than performance and this suggested relationship to behaviour is the basis of many studies looking at virtual reality for treatment with exposure therapy. This area of work is currently being investigated by a number of groups and has had very good success to date, for example when being used to treat phobias, where patients learn how to control their automatic responses to anxiety-provoking situations. Here the response, such as fear and heart rate change is according to Herbelin, Vexo & Thalmann (2002) more closely linked to being emotionally involved and effective rather than if the system is realistic. Studies in this area therefore support the proposed link between presence and behaviour.

This lack of relationship between presence and performance disagrees with the earlier published work by Witmer & Singer (1998) who found a positive significant relationship between presence and task performance in one experiment, and a positive but not significant relationship in a second experiment. However, they argued that it should exist and that the reason it was present, but not found to be significant in the second experiment was due to both performance and presence being dependent on many different factors, such as an individual's skills and abilities.

The finding of this study suggests that performance might be more closely linked to immersion and that presence may only have a small effect on performance. As stated earlier in the discussion in section 7.3 this might be due to the fact that immersion is but one component making up presence, with many other factors also contributing to the perception of presence, such as length of exposure, personality and social factors which might be more important when looked at together compared with immersion. This finding and supposition is supported by Slater *et al*. (1996) who found that during a simulated chess game task performance was related to immersion, but not presence. However, it must be pointed out that there are differences in Slater's study and the one performed here. Slater *et al*. (1996) were looking at a game rather than the real life scenario of clinical competency and also compared the use of a head mounted virtual display to viewing and interacting with the game on a TV screen.

This finding has potential impact on the design and use of VR simulation within education and in particular medical and para-medical simulation as it is suggests that unlike training and transfer of knowledge that has been linked to presence, for the assessment of competency presence might be less important a consideration and that immersion might become the important issue to consider when designing VR systems for competency measurement.

However, this also brings in to question what we want from a competency measure. If you are trying to measure competency in VR simulators with low presence the user will not feel the psychological pressure that comes with the situation, and so may not react as they would in real life. They may pass the competency, but not be able to perform the same actions in real life when other stressors are present. With a highly immersive simulation one could create an environment that elicits a feeling of high presence in the user; in this way the user will perform and react as they might in real life. This situation, which could possibly be only achieved with VR might allow medics and para-medical trainers not only to measure a base competence, but also look at a competence under different levels of stress in a controlled way. It may therefore be possible to look at competencies in high presence environments exposing the user to different levels of stress to see if the user maintains a stable performance. This would then allow educators to measure performance and using assessment feedback, to look at and modify the actions of the trainees for a given competency; but it would also allow an opportunity to look at and modify the behavioural responses of the users for the same competence.

7.8.6 Content Validity

As with part of the quantitative analysis, aspects of the qualitative analysis can be considered as a study on validity, in this case content validity. Subjects rejected the use of the VERT simulation as an assessment tool commenting extensively on why this is the case and so this use of VERT lacks content validity. Students commented more favourably on using VERT as a learning tool remarking it was "useful", "it does help, the visual learning", "you also get used to using the hand pendant". As a learning tool VERT is far more acceptable to the subjects, particularly when looking at its use early on in the course, in fact some focus groups considered the practice of electron setups to be a good use of the VERT system so indicating that content validity might not be depended only on the simulation but the use the simulation is put to.

Although there were some negative comments about the use of the LINAC this was generally seen in a more favourable light by students who felt that it was more realistic in a number of ways and did comment that it could be used as a measure of competence so lending it some degree of content validity as an assessment tool.

7.8.7 Final thoughts and Chapter Summary

When looking at the results from the qualitative study it is striking how well the students' narratives from the focus groups corroborate and support the findings from the statistical analysis. The quantitative part of the analysis clearly identifies performance on the LINAC simulation is consistently better than that on VERT for virtually all components of the setup. In the qualitative findings students explain issues with VERT that may explain issues they felt they had in both the VERT and the simulations, but the main issues were with VERT.

Both strands of the work show in a convincing way how VERT is currently unsuitable for use as an assessment tool for measuring competency for electron setups. What is apparent however, from the focus groups is the student's lack of perception about learning from assessment. The students appear to have only considered the assessment as a means of measuring their ability to perform the setups and not consider what they have learnt from doing the assessment. Students were seen to adapt their practice and attempted to find a work around to the differences they encountered. Although as stated earlier it is not usual to use lasers to routinely set FSD some students did to overcome the issue in not being able to see the FSD light. Some students did stand in or look though solid objects on VERT in order to overcome issues with restricted views. This lack of understanding about what assessments are for and the apparent thought that it was the end point assessment rather than a learning opportunity may have coloured their views about the suitability of VERT as an assessment tool especially as the students felt that VERT does have a role to play in preparing them for assessment.

Although some students in the focus groups felt that the order they did the simulations in affected their performance the data analysis did not show this to be the case with all findings. There were order effects for the first electron setup, but on no other timings and none of the competency measures, so the results were largely independent of the sequence they were performed and the washout period had had the desired effect.

Looking at the results in terms of content validity the findings could be reworded stating that content validity is lacking for the VERT assessment, but is present for the LINAC simulation.

In the final concluding chapter the pertinent points from the preceding chapters will be summarised. There will also be recommendations on how to take the findings of this study forward.

Chapter 8 - Key findings and implications for research

The main aims of this research were two-fold. First, to compare the use of two as near identical electron setup simulations, one virtual and one clinically based to measure competency. Secondly, to find how acceptable VERT was as a measure of competency to the students themselves.

The two simulations gave very different results in performance; students performing significantly better in the LINAC simulation compared to the VERT simulation in virtually all of the setup parameters. The focus groups also found a number of key factors that the students felt affected their performance. The main factor that applied to both simulations was that of fidelity, or more specifically lack of fidelity especially in relation to the phantom not being a whole patient phantom. This finding fits in with the model proposed by Aggarwal *et al.* (2010) that high fidelity simulations are best in advanced situations with experienced learners.

This thought can be extended further when considering the individual measures that made up the overall score. Differences in many of the individual measures could again be linked back to the lack of physical fidelity on VERT compared with the LINAC simulation such as inability to see the FSD light clearly or interact with the couch controls as well as issues with the environmental fidelity (smells and sound) and psychological fidelity "*it didn't feel right*". Additionally it can also be seen from the students' use of the machine movements that there were slight differences in the procedural movements during the first and final 30 seconds of the treatment setup which might be taken as a difference in cognitive fidelity. Fidelity is therefore a key factor in the student's performance and should be a major consideration in the development and design of future simulations that are intended to be used as tools for assessment. Lack of fidelity impacts directly with the outcome and also affects the subjects procedural selection during the process making the subjects choose alternative pathways to try and achieve the same outcome.

One cannot with 100% conviction state which performance mirrored best what the performance in real life would be, but it is likely that the LINAC performance was closest given that the only difference in the setup between this and real life was that a phantom replaced the patient. The focus group feedback which identified issues with fidelity, environment, assessment and the equipment, most of which did not relate to the LINAC simulation would also suggest that the students considered the LINAC more suitable as a measure of competence than VERT. Also the presence felt by the subjects on the LINAC was substantially higher than that for the VERT simulation, which is generally considered to be a favourable point. These aspects lead to the inference that the VERT simulation lacks construct validity.

A number of hypotheses were put forward as to the reason behind the observed differences based on the information from the focus groups. These included a lack of contextualised learning, stress and incorrect sensory feedback leading to a loss of spatial ability which in turn can be associated with lack of verity of the VERT simulation, environmental factors, equipment controls not all being represented and a generally lower sense of presence compared to the LINAC simulation. These elements either in a combined\manner or individually produced differences in the use of the two simulations in the final 30 seconds of use.

Presence was not found to be significantly related to competency scores in either simulation. This might be because immersion is more important factor to consider as it appears that presence may be more closely related to behaviour rather than performance. Immersion was not directly measured in this study, but many of the points raised by the students in the focus groups would support this finding.

When looking at elements of the results as issues of validity it was clear that the VERT simulation as a tool for measuring competency lacked both content and construct validity and did not have concurrent validity with the LINAC simulation. The LINAC simulation was better regarded by the students and could be said to have content validity although this could be improved further with modifications to the simulation particularly in relation to the patient. In terms of construct validity the LINAC simulation did not show any significant differences between the three cohorts, but there was a tendency for the score to increase with each cohort's experience and there is at least an indication that with a better designed study as regards each cohort's sample size this finding might be improved upon.

What was also clear was that not only were there differences in outcome between the two simulations, but also the way students acted within the simulations was also different. Although automatic movements were similar other movements made by students varied considerably. Also students undertook more multiple movements on the LINAC simulation compared to VERT. Another difference was the amount of inactivity which was significantly higher on VERT. Differences in the verity and realism of the simulation therefore affects the user's performance and although some of this difference could be attributed to lack of couch controls, not all could be.

Despite issues in using the VERT simulation as an assessment tool subjects felt that the system afforded them an opportunity to experience and learn on the simulator. Their negativity to VERT as an assessment tool came from a wide range of issues and despite the difference in setup outcomes students still felt that VERT offered substantial learning opportunities and the freedom to learn experientially without time pressures, the lack of reality not being considered as important for learning as it was in assessment. Findings from the qualitative data appear to indicate that content validity exists for the LINAC simulation, but not for the VERT simulation when looking to assess the competency of the students, but probably exists for both when used as a learning tool. Consideration of the intended use the simulator is being put to is therefore as important as the simulator and what it offers the users when considering its validity. It cannot be assumed that simulators that have published validity for training are valid for other uses, and a separate validity study would need to be undertaken depending on what the simulator is being used for, basic training, advanced training, formative or summative assessment.

8.1 Limitations and Boundaries

Maintaining the rigor in any research project is critical; however, this need must be balanced by the inherent practical barriers that arise when undertaking the research so that there is a balance between rigor and practicality. All research will therefore have limitations that must be noted.

Firstly, both the quantitative and especially the qualitative aspects of this study have the limitation of a lack of generalisability. Results are derived from a small population and expansion of these results and findings to other populations beyond which this research focuses on may not be acceptable. This is not to say that the sample size of the study was small, but rather acknowledges two points. Firstly that VR simulators exist in different forms and the results found here that apply to VERT and therapeutic radiographers may not all apply to other VR simulations. Secondly to acknowledge that this was a single centre study, which like all studies has strengths as well as weaknesses. Single centre studies are logistically easier, they simplify data collection and they typically have a more homogenous population. Certainly in this instance all the students in each cohort had been exposed to VERT in the same way and had all had the same training within their respective cohort. Unfortunately this last point is also a disadvantage in that because of this homogeneity the generalisability of the study to the population as a whole is not as good and there is potentially limited external validity, but as a first study in this area the study is still very useful.

Replication of this study utilising different simulations and simulators in multicentre trials is recommended as future research to determine whether the identified pattern of relationships is restricted to the present design or student group.

Finally the design of the study precludes a definitive answer of which simulation estimates/replicates a student's performance in reality. Certainly from the student's responses in the focus groups it likely that the LINAC simulation was more similar to the student's expected performance. Equally, it is an unlikely supposition that students constantly over performed on the LINAC compared to how they would perform with a real patient with the additional time pressures. These two arguments would therefore suggest that the performance on the LINAC was closer to reality and therefore the VERT results were further from the true performance of the students. The probable reasons for this have been mentioned above, but this is supposition and would need to be investigated further in future research.

8.2 Future Directions

The current study suggests a number of limitations both with the system and the way the simulations were undertaken. The results from this study have been fed back to Vertual, the manufacturers of the VERT system and a number of issues raised by this study are in the process of being addressed.

These findings contribute to our knowledge of virtual reality in two ways. Firstly consideration needs to be given both to the student's performance and their expectations of the simulators in a number of different scenarios. This needs to be done for both live and virtual simulations in order to ensure that both learning and assessment conditions are valid. Authors need to move away from stating in publications that the simulator has proven validity and consider more that, the simulation undertaken has validity. If the simulation or use of the simulator changes, then so might the validity. Evaluation of other simulations are therefore needed comparing their use, both as a learning tool and as a measure of assessment. Secondly, that competency assessment requires a higher verity and reality in the simulation than does student learning. Further research could be done to consider this aspect looking at assessment clarifying the students' views regarding the use VR for assessment for learning and assessment of learning.

Building on this thought the work detailed in this thesis has led to the current programme content being updated and VERT has been introduced into the programme in the assessment role, not as a summative measure of assessment, but rather a formative assessment tool. Students are allowed to practice an electron setup and then asked to undertake an observed assessment. The assessment is timed and at the end of the assessment there is a student debrief, where the final setup is reviewed and discussed with the student about the decisions identified by the observer during the setup. The score derived from the standoff and time is then given to the student. The score is then written on a magnetic strip and placed on a board. A formal evaluation of this is still needed, but student feedback through the module evaluation route about this change has been very supportive of its introduction.

There is a lack of multi centre trials in this area; this study and all the studies in Appendix XI are single centre trials. Larger trials are needed to ensure findings are applicable to the population as a whole. Doing this should also allow multi-level modelling to be used in the analysis which would allow possible interaction effects such as that that might exist between age and cohort to be analysed.

One unexpected finding emerged in relation to Aim 2 where it was found that there was a significantly higher presence score in the real environment compared to the virtual environment. This was contrary to the findings of Usoh *et al*., (2000) which found no significant difference and therefore concluded that the questionnaire was unsuitable for cross environment comparisons. Because of this contradiction with the only other published research on the use of the PQ in real life this area also warrants further investigation.

Finally no significant relationship could be found between presence and performance during the assessment as measured by the competency scores. It was suggested in the text that immersion might be a more important concept than presence when looking at performance and this relationship should be explored further.

References

Aggarwal, R., Crochet, P., Dias, A., Misra, A., Ziprin, P., Darzi, A. (2009) “Development of a virtual reality training curriculum for laparoscopic cholecystectomy.” *British Journal of Surgery*. 96: 1086-1093.

Aggarwal, R., Grantcharov, T., Moorthy, K., Milland, T., Papasavas, P., Dosis, A., Bello, F., Darzi, A. (2007) “An Evaluation of the Feasibility, Validity, and Reliability of Laparoscopic Skills Assessment in the Operating Room.” *Annals of Surgery*. 245: 992-999.

Aggarwal, R., Mytton, O., Derbrew, M., Hananel, D., Heydenberg, M., Issenberg, B., MacAulay, C., Mancicn, M. Morimoto, T. Soper, N. Ziv, A. and Reznick, R. (2010) “Training and simulation for patient safety.” *Quality & Safety in Health Care*. 19: i34-i43

Albrecht, J., Meves, A., Bigby, M. (2005) “Case reports and case series from Lancet had significant impact on medical literature.” *Journal of Clinical Epidemiology*. 58: (12), 1227-1232.

Albrecht, K., Blom, K.J., Beckhaus, S. (2009) “Perceived Affordances: Why do people wear virtual cooking pots on their heads?” [Proceedings of the *International Conference on Presence*.] Los Angeles.

Alessi, S.M., Trollip, S.R. (2001) “Multimedia for Learning: Methods and Development.” (3rd Edition) Allysn and Bacon Inc. Boston.

Alessi, S. (2000) “Simulation design for training and assessment.” In: O’Neil, H.F., Andrews, D.H. (Editors) Aircrew Training and Assessment. Lawrence Erlbaum Associates Publishers.

Alinier, G. (2007) “A typology of educationally focused medical simulation tools.” *Medical Teacher*. 29: e243–e250.

Alsina-Jurnet, I., Beciu, C.C., Maldonado, J.G. (2005) "Individual Differences in the Sense of Presence." *Presence.* [Paper presented at the Eighth Annual International Workshop on Presence, London, UK.] 133-142.

Alsina-Jurnet, I., Gutiérrez-Maldonado, J. (2010) "Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments." *International Journal of Human-Computer Studies*. 68: 788–801.

ARRT (2012) "Radiography Didactic and Clinical Competency Requirements." ARRT. St Paul.

ARRT (2013) "Radiation Therapy Didactic and Clinical Competency Requirements." ARRT. St Paul.

Arunkumar, T., Supe, S. S., Ravikumar, M., Sathiyan, S., Ganesh, M. (2010) "Electron beam characteristics at extended source-to-surface distances for irregular cut-outs." *Journal of Medical Physics.* 35: (4), 207-214.

Ashcraft, M.H., Radvansky, G.A. (2014). Cognition. 5th Edition. Pearson.

Australian Institute of Radiography (2013) "Comments on the content of the draft capabilities document on Professional capabilities for medical radiation practice." Darcy Associates Consulting Services.

Bagshaw, S.M., Bellomo, R. (2008) "The need to reform our assessment of evidence from clinical trials: A commentary." Philosophy, Ethics, and Humanities in Medicine. 3: (23), 1-11.

Bajka, M., Tuchschmid, S., Fink, D., Székely, G., Harders, M. (2009) "Establishing construct validity of a virtual-reality training simulator for hysteroscopy via a multimetric scoring system." *Surgical Endoscopy*. [Downloaded April 2015]
https://www.vision.ee.ethz.ch/publications/papers/articles/eth_biwi_00689.pdf

Barrett, A., Dobbs, J., Roques, T. (2009) “Practical Radiotherapy Planning.” 4th Edition. CRC Press.

Bartell, T.G., Johnson, K.R. (2013) “Making Unseen Privilege Visible in Mathematics Education Research.” *Journal of Urban Mathematics Education.* 6: (1), 35–44.

Bearman, M., Nestel, D., Andreatta, P. (2013) “Simulation-based medical education.” In: Walsh, K. (Editor) The Oxford Book of Medical Education. Oxford University Press, Oxford.

Bedny, G.Z., Karwowski, W., Bedny I.S. (2012) “Complexity Evaluation of Computer-Based Tasks.” *International Journal of Human-Computer Interaction*. 28: (4), 236-257.

Bell, J.T., Fogler, H.S. (1995) “The Investigation and Application of Virtual Reality as an Educational Tool.” [Proceedings of the American Society for Engineering Education Annual Conference.] Anaheim, CA.

Bittner, J.G., Mellinger, J.D., Imam, T., Schade, R.R., MacFadyen, B.V. (2010) “Face and construct validity of a computer-based virtual reality simulator for ERCP.” *Gastrointestinal Endoscopy*. 71: (2), 357-364.

Bland, J.M., Altman, D.G. (1986) “Statistical methods for assessing agreement between two methods of clinical measurement.” *Lancet*, i: 307-310.

Blaufox, M.D. (1996) “Becquerel and the discovery of radioactivity: early concepts.” *Seminars in nuclear medicine*. 26: (3), 145-54.

Blom, K.J. (2010) “Virtual Affordances: pliable user expectations.” In Steinicke, F., Willemsen, P. (Editors) 2nd IEEE VR Workshop on Perceptual Illusions in Virtual Environments.

Bohm, D. (1980) “Wholeness and the Implicate Order.” Routledge; Reissue Edition (2002).

Bongers, P.J., van Hove, P.D., Stassen, L.P.S., Dankelman, J., Schreuder, H.W.R. (2014) "A New Virtual-Reality Training Module for Laparoscopic Surgical Skills and Equipment Handling: Can Multi-tasking be Trained? A Randomized Controlled Trial." *Journal of Surgical Education*. 72: (2), 184-191.

Boud, D. 2010. "*Assessment for Developing Practice*." In Higgs, J., Fish,D., Goulter, I., Loftus, S., Reid, J.A., Trede, F. (Editors) Education for Future Practice. 251-262. Rotterdam: Sense.

Boud, D., Soler, R. (2015) "Sustainable assessment revisited" *Assessment & Evaluation in Higher Education*. [Downloaded June 2015] http://dx.doi.org/10.1080/02602938.2015.1018133

Bower, M., Howe, C., McCredie, N., Robinson, A., Grover, D. (2014) "Augmented Reality in education - cases, places and potentials." *Educational Media International*. 51: (1), 1-15.

Bowman. D.A., McMahan, R.P. (2007) "Virtual Reality: How much Immersion is enough?" *Computer*. 40: (70), 36-42.

Brady, S., Bogossiana, F., Gibbons, K. (2015) "The effectiveness of varied levels of simulation fidelity on integrated performance of technical skills in midwifery students - A randomised intervention trial." *Nurse Education Today*. 35: (3), 524-529.

Brannen, J. (2005) "Mixed Methods Research: A discussion paper". NCRM Methods Review Papers, NCRM/005. [Downloaded January 2014] http://eprints.ncrm.ac.uk/89/

Brannen, J. Halcomb, E.J. (2009). "Data collection in mixed methods research." In Andrew, S., Halcomb, E.J. (Editors) Mixed Methods Research for Nursing and the Health Sciences (pp. 67-83). Chichester: Wiley-Blackwell.

Brecher, R., Brecher, E. (1969) "The Rays: a History of Radiology in the United States and Canada." Baltimore: Williams and Wilkins.

Brewin, J., Ahmed, K., Khan, M.S., Jaye, P., Dasgupta, P. (2014) "Face, Content, and Construct Validation of the Bristol TURP Trainer." *The Journal of Surgery*. 71: 500-505.

Bright, E., Vine, S., Wilson, M.R., Masters, R.S.W., McGrath, J.S. (2012) "Face validity, construct validity and training benefits of a virtual reality TURP simulator." *International Journal of Surgery*. 10: 163-166.

Brown, J.D. (2000) "Questions and answers about language testing statistics: What is construct validity?" JALT Testing & Evaluation SIG Newsletter. 4: (2), 8-12.

Brown, M.J., Carlson, D.J., Brenner, D.J. (2014) "The Tumor Radiobiology of SRS and SBRT: Are More Than the 5 Rs Involved?" *International Journal of Radiation Oncology, Biology and Physics*. 88: (2), 254–262.

Brown, T.H. (2005) "Beyond constructivism: Exploring future learning paradigms." *Education Today*. 2. Aries Publishing Company, Thames, New Zealand.

Brydges, R., Hatala, R., Zendejas, B., Erwin, P.J., Cook. (2015) "Linking simulation-based educational assessments and patient-related outcomes: A systematic review and meta-analysis." *Academic Medicine*. 90: 246-256.

Bryman, A. (2007) "Barriers to Integrating Quantitative and Qualitative Research." *Journal of Mixed Methods Research*. 1: (1), 8-22.

Bulu, S.T. (2012) "Place presence, social presence, co-presence, and satisfaction in virtual worlds." *Computers & Education.* 58: 154-161.

Bystrom, K.E., Barfield, W., Hendrix, C. (1999). "A conceptual model of sense of presence in virtual environments." *Presence: Teleoperators and Virtual Environments*. 5: (1), 109-121.

Cameron, R. (2011) "Mixed Methods Research: The Five Ps Framework." The Electronic *Journal of Business Research Methods*. 9: (2), 96-108. [Downloaded December 2013] www.ejbrm.com

Cancer Reform Strategy (2007) Department of Health.

Cant, R. P., Cooper, S. J. (2010) "Simulation-based learning in nurse education: systematic review." *Journal of Advanced Nursing*. 66: (1), 3-15.

Carless, D. (2009) "Trust, distrust and their impact on assessment reform." *Assessment & Evaluation in Higher Education*. 34, (1), 79-89.

Carr, T., Cox, L., Eden, N., Hanslo, M. (2004) "From peripheral to full participation in a blended trade bargaining simulation." *British Journal of Educational Technology*. 35: (2), 197-211.

Castle, A., Holloway, D. and Rage, A. (1998) "A review of issues in continuing professional development in teaching, nursing and radiography." *International Journal of Lifelong Education* 17: (5), 328-340.

Chandler, D. (1995) "Technological or Media Determinism." URL [Downloaded December
2013] http://www.aber.ac.uk/media/Documents/tecdet/tecdet.html

Chavez, C. (2008) "Conceptualizing from the Inside: Advantages, Complications, and Demands on Insider Positionality." Qualitative Report. 13: (3), 474-494.

Chemero, A. (2003) "An Outline of a Theory of Affordances." *Ecological Psychology*. 15: *(*2), 181-195.

Chopra, V., Gesink, B.J., DeJong, J., Bovill, J.G., Spierdijk, J., Brand, R. (1994) "Does training on an anaesthesia simulator lead to improvement in performance?" *British Journal of Anaesthesia.* 73: (3), 293-297.

Cicchetti, D. V., Shoinralter, D., Tyrer, P. J. (1985) "The Effect of Number of Rating Scale Categories on Levels of Interrater Reliability: A Monte Carlo Investigation." *Applied Psychological Measurement*. 9: (1), 31-36.

Coombs, S.J. (2003) "The Hawthorne Effect: Is it a help or hindrance in social science research?" *Change: Transformations in Research*. 6: (1), 97-111.

Cooper, J.B., Taqueti, V.R. (2004) "A brief history of the development of mannequin simulators for clinical education and training." *Quality and Safety in Health Care*. 13: (Suppl 1), i11-i18.

Cooper, S., Cant, R., Porter, J., Bogossian, F., McKenna, L., Brady, S., Fox-Young, S. (2012) "Simulation based learning in midwifery education: A systematic review." *Women and Birth*. 25: 64-78.

Costello, P.J. (1997) "Health and Safety Issues associated with Virtual Reality - A Review of Current Literature." Advisory Group on Computer Graphics (AGOCG). Technical Reports.

Cox, M.J. (2013) "Formal to informal learning with IT: research challenges and issues for e-learning." *Journal of Computer Assisted Learning*. 29: (1), 85-105.

Creswell, J.W. (2002) "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches". 2nd Edition SAGE Publications, Inc.

Creswell, J.W., Plano Clark V.L. (2010) "Designing and Conducting Mixed Methods Research." SAGE Publications, Inc; 2nd Edition.

Cruickshank, J. (2012) "Positioning positivism, critical realism and social constructionism in the health sciences: a philosophical orientation". *Nursing inquiry*. 19: (1), 71-82.

Curran, V.R., Aziz, K., Siu O'Young, S., Bessell, C. (2010) "Evaluation of the Effect of a Computerized Training Simulator (ANAKIN) on the Retention of Neonatal Resuscitation Skills." Teaching and Learning in Medicine: An International Journal. [Downloaded January 2015] http://www.tandfonline.com/loi/htlm20

Dabrowski, I.J. (1995) “David Bohm's Theory of the Implicate Order: Implications for Holistic Thought Processes.” *Issues in Integrative Studies*. 13: 1-23.

Dalgarno, B., Lee, M.J.W. (2012). “Exploring the relationship between afforded learning tasks and learning benefits in 3D virtual learning environments.” In *Future challenges, sustainable futures.* Proceedings ascilite Wellington,

Dalziel, K., Round, A., Stein, K., Garside, R., Castelnuovo, E., Payne, L. (2005) “Do the findings of case series studies vary significantly according to methodological characteristics?” Health Technology Assessment. 9: (2).

Datta, R., Upadhyay, K.K., Jaideep, C.N. (2012) “Simulation and its role in medical education.” *Medical Journal Armed Forces India*. 68: (2), 167-172.

Davies, P., Gribbin, J. (1992). “The matter myth.” New York: Simon & Schuster.

Debling, G. (1992) “Building Quality Assurance into Local Assessment: The need to be cost effective.” *Competence and Assessment: Compendium* 2, Employment Department, Sheffield, UK.

De Jong, T., van Joolingen, W.R. (1998) “Scientific Discovery Learning with Computer Simulations of Conceptual Domains.” *Review of Educational Research*. 68: (2), 179-201.

de Kort, Y.A.W., IJsselsteijn, W.A., Kooijman, J., Schuurmans, Y. (2003) “Virtual Laboratories: Comparability of Real and Virtual Environments for Environmental Psychology.” *Presence*. 12: (4), 360–373.

de Kraker, J., Lansu, A., van Dam-Mieras, M.C. (2007) “Competences and competence-based learning for sustainable development.” In de Kraker, J., Lansu, A., van Dam-Mieras, M.C. (Editors) Crossing boundaries. Innovative learning for sustainable development in higher education. (103-114). Frankfurt am Main:Verlag für Akademische Schriften.

Delaney, G., Jacob, S., Featherstone, C., Barton, M. (2005) "The Role of Radiotherapy in Cancer Treatment. Estimating Optimal Utilization from a Review of Evidence-Based Clinical Guidelines." *Cancer*. 104: 112937.

DHSS (1979) Letter from the DHSS to Michael Jordan, Secretary, 18th June 1979. College of Radiographers Archive.

DoH (2003) "Radiography skills mix: a report on the four-tier service delivery model." Department of Health.

Downing, S. M., Haladyna, T. M. (2004) "Validity threats: overcoming interference with proposed interpretations of assessment data." *Medical Education*. 38: 327–333.

Draper, J.V., Kaber, D.B., Usher, J.M. (1998) "Telepresence." *Human Factors*. 40: (3), 354-375.

Eichenbaum, H. (2013) "Memory on time." Trends in Cognitive Sciences. 17: (2), 81-88.

Elo, S., Kyngäs, H. (2008) "The qualitative content analysis process." *Advanced Nursing*. 62: (1), 107-115.

Eraut, M. (1994) "Developing Professional Knowledge and Competence." The Farmer Press. London.

Eriksen, J.R., Grantcharov, T. (2005) "Objective assessment of laparoscopic skills using a virtual reality stimulator." *Surgical Endoscopy*. 19: 1216-1219.

Faas, D., Bao, O., Frey, D.D., Yang, M.C. (2014) "The influence of immersion and presence in early stage engineering designing and building." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. 28: (2), 139-151.

Felder, R.M., Silverman, L.K. (1988). "Learning and teaching styles in engineering education." *Engineering. Education.* 78(7), 674-681. [Downloaded August 2013] http://www4.ncsu.edu./unity/lockers/users/f/felder/public/Papers/LS- 1988.pdf

Felnhofer, A., Kothgassner, O.D., Beutl, L., Hlavacs, H., Kryspin-Exner, I. (2012) "Is Virtual Reality made for Men only? Exploring Gender Differences in the Sense of Presence." Paper in Conference Proceedings or in Workshop Proceedings (Paper). International Society for Presence Research Annual Conference – ISPR.

Feng, J., Spence, I., Pratt, J. (2007) "Playing an Action Video Game Reduces Gender Differences in Spatial Cognition." *Psychological Science*. 18: (10), 850-855.

Fernald, D.H., Coombs, L., DeAlleaume, L., West, D., Parnes, B. (2012) "An Assessment of the Hawthorne Effect in Practice-based Research." *Journal of the American Board of Family Medicine*. 25: (1), 83-86.

Fetscherin, M., Lattemann, C. (2008) "User Acceptance of Virtual Worlds." *Journal of Electronic Commerce Research*. 9: 3, 231-242.

Fielding, K., Shapiro, R., Tang, E. (2012) "Responding to an emergency – how stress affects medical trainees." *University Of Western Ontario Medical Journal*. 81: (1), 39-40.

Flinton, D. (2009) in Cherry, P., Duxbury, A. M. "Practical Radiotherapy. Physics and Equipment." 2nd Edition. Wiley Blackwell.

Flinton, D., Miles, E. (2009) in Cherry, P., Duxbury, A. M. "Practical Radiotherapy. Physics and Equipment." 2nd Edition. Wiley Blackwell.

Flinton, D., White, N. (2009) "Preliminary findings on the Virtual Environment for Radiotherapy Training (VERT) system: simulator sickness and presence." *Journal of Radiotherapy in Practice*. 8: (4) 169-176.

Floridi, L. (2005) "The Philosophy of Presence: From Epistemic Failure to Successful Observation." *Presence: Teleoperators and Virtual Environments*. 14: (6), 656-667.

Fogg, B.J. (2002) “Persuasive Technology: Using Computers to Change What We Think and Do (Interactive Technologies).” Morgan Kaufmann Publishers Inc.

Fowler, J.F. (2010) “21 years of Biologically Effective Dose.” *The British Journal of Radiology*. 83: (991), 554-568.

Gaba, D. M. (2004) “The future vision of simulation in health care.” *Quality & Safety in Health Care*. 13: i2-i10.

Gardner, J. (2012) “Assessment and Learning.” 2nd Edition. Sage Publications Ltd.

Garson, G. D. (2009). “Computerized simulation in the social sciences.” *Simulation & Gaming*. 40: (2), 267-279.

Gavazzi, A., Bahsoun, A.N., Van Haute, W., Ahmed, K., Elhage, O., Jaye, P., Khan, M.S., Dasgupta, P. (2011) “Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot).” *Annals of the Royal College of Surgeons of England*. 93: (2), 146–150.

Gaver, W.W. (1991). “Technology affordances.” CHI ’91: Proc. SIGCHI Conference on Human Factors in Computing Systems, 79-84. New York: ACM.

Gibbs, G., Simpson, C. (2004) “Conditions under which assessment supports students’ learning’. Learning and Teaching in Higher Education. 1, 3-31.

Gibson, J.J. (1986) “The Ecological Approach to Visual Perception.” Psychology Press; Taylor Francis.

Gilakjani, A.P., Leong, L.M., Ismail, H.N. (2013) “Teachers’ Use of Technology and Constructivism.” *International Journal Modern Education and Computer Science*. 4: 49-63.

Glavin, R.J., Maran, N.J. (2002) “Development and use of scoring systems for assessment of clinical competence.” *British Journal of Anaesthesia*. 88: (3), 329-30.

Gonczi, A. (1993) “Applying Competency Standards in the Professions.” Paper presented at the *Continuing Professional Education Conference: A question of competence,* May 1993, Coffs Harbour NSW.

Goodie, A.S., Crooks, C.L. (2004) “Time - Pressure Effects on Performance in a Base - Rate Task.” *The Journal of General Psychology*.131: (1), 18-28.

Grady, J.L., Kehrer, R.G., Trusty, C.E., Entin, E.B., Entin, E.E., Brunye, T.T. (2008) “Learning Nursing Procedures: The Influence of Simulator Fidelity and Student Gender on Teaching Effectiveness.” *Journal of Nursing Education*. 47: (9), 403-408.

Grantcharov, T.P., Bardram, L., Funch-Jensen, P., Rosenberg, J. (2003) “Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills.” *The American Journal of Surgery*. 185: 146-149.

Gredler, M.E. (2004) “Games and simulations and their relationships to learning.” In Jonassen, D.H. (Editor) Handbook of research for educational communications and technology, 2nd Edition. Lawrence Erlbaum Associates. Mahwah, NJ.

Green, J., Thorogood, N. (2014) “Qualitative Methods for Health Research.” 3rd Edition. Sage Publications Ltd.

Griffiths, S., Short, C. (1994) “Radiotherapy: Principles to Practice.” Churchill Livingstone. Edinburgh.

Grimshaw, M., Charlton, J.P., Jagger, R. (2011) “First-Person Shooters: Immersion and Attention.” *Journal for Computer Game Culture*. 5: (1), 29-44.

Grosas, A.B., Raju, S.R., Schuett, B.S., Chuck, J., Millar, T.M. (2014) “Determining if active learning through a formative assessment process translates to better performance in summative assessment.” *Studies in Higher Education*. [Downloaded August 2014] http://dx.doi.org/10.1080/03075079.2014.988704

Gunkel, D.J. (2003). “Second thoughts: toward a critique of the digital divide.” *New Media and Society*. 5: (4), 499–522.

Halkett, G.K.B., Kristjanson, L.J., Lobb, E., Little, J., Shaw, T., Taylor, M., Sprye, N. (2012) “Information needs and preferences of women as they proceed through radiotherapy for breast cancer.” *Patient Education and Counseling*. 86: (3), 396-404.

Hall, R. (2012) “Mixed methods: In search of a paradigm.” Proceedings from Phuket 2012: International Conference: Innovative Research in a Changing and Challenging World. [Downloaded May 2013] http://auamii.com/proceedings_phuket_2012/hall.pdf

Hammond, M. (2010) “What is an affordance and can it help us understand the use of ICT in education?” *Education and Information Technologies.* 15: 205-217.

Hartson, H.R. (2003) “Cognitive, sensory, and functional affordances in interaction design.” *Behaviour and Information Technology*. 22: (5), 315 - 338.

Harvey, A., Bandiera, G., Nathens, A., LeBlanc, V.R. (2012) “Impact of stress on resident performance in simulated trauma scenarios.” *Journal of Trauma and Acute Care Surgery*. 72: (2), 497-503.

Haycock, A. (2011) “Moving from training to competency testing.” *Techniques in Gastrointestinal Endoscopy*. 13: (2), 155–160.

Hennink, M.M., (2014) “Understanding Focus Group Discussions.” Oxford University Press.

Henry, A.D., Baron, K.B., Mouradian, L., Curtin, C. (1999). “Reliability and validity of the Self-Assessment of Occupational Functioning.” *American Journal of Occupational Therapy*. 53: 482-488.

Henwood, S.M., Flinton, D. (2012) “5 years on: Have attitudes towards continuing professional development in radiography changed?” *Radiography*. 18: (3), 179-183.

Henwood, S.M., Yielder, J., Flinton, D. (2004) "Radiographers attitudes to mandatory CPD: a comparative study in the United Kingdom and New Zealand." *Radiography*. 10: (4), 251-258.

Herbelin, B., Vexo, F., Thalmann, D. (2002) "Sense of Presence in Virtual Reality Exposures Therapy." In 1st. International Workshop on Virtual Reality Rehabilitation.

Herer, A., Galloway, R.A., Cleland, M.R., Berejka, A.J., Montoney, D., Dispenza, D., Driscoll, M. (2009) "X-ray-cured carbon–fiber composites for vehicle use." *Radiation Physics and Chemistry*. 78: 531-534.

Hill, T., Lewicki, P. (2005) "Statistics Methods and Applications." Statsoft Incorporated.

Hitchcock, G., Hughes, D. (1995) "Research and the Teacher: A Qualitative Introduction to School-based Research." Psychology Press.

Hodges, L.F., Rothbaum, B.O., Kooper, R., Opdyke, D., Meyer, T., de Graaff, J.J., Williford, J.S., North, M.M. (1994) "Presence as the defining factor in a VR application. Virtual Reality Graded Exposure in the Treatment of Acrophobia." Technical Report GVU-94-06. Georgia Tech University, Graphics, Visualization, and Usability Center.

Hogle, N.J., Briggs, W.M., Fowler, D.L. (2007) "Documenting a Learning Curve and Test-Retest Reliability of Two Tasks on a Virtual Reality Training Simulator in Laparoscopic Surgery." *Journal of Surgical Education*. 64: 424-430.

Holloway, I., Wheeler, S. (2009) "Qualitative Research in Nursing and Health Care." 3rd Edition. Wiley-Blackwell.

Howe, K.R. (2012) "Mixed Methods, Triangulation, and Causal Explanation." *Journal of Mixed Methods Research*. 6: (2), 89-96.

Howell, D.C. (2008) "The analysis of missing data." In Outhwaite, W., Turner, S. Handbook of Social Science Methodology. London, Sage.

Huang, H.M., Liaw, S.S., Lai, C.M. (2013) "Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems." *Interactive Learning Environments*. [Downloaded December 2013] http://www.tandfonline.com/doi/abs/10.1080/10494820.2013.817436#.VM6Dey493Rg

Hunt, R. J. (1986) "Percent Agreement, Pearson's Correlation, and Kappa as Measures of Inter-examiner Reliability." *Journal of Dental Research*. 65: (2), 128-130.

Insko, B.E. (2003) "Measuring Presence: Subjective, Behavioral and Physiological Methods." In Riva, G., Davide, F., IJsselsteijn, W.A. (Editors) Being There: Concepts, effects and measurement of user presence in synthetic environments. IOS Press.

International Atomic Energy Agency (2014) "A Handbook for the Education of Radiation Therapists (RTTs)." IAEA, Vienna.

Ioanna, I., Zhou, Y., Szudek, J., Kennedy, G., O'Leary, S. (2014) "The Effect of Fidelity: How Expert Behavior Changes in a Virtual Reality Environment." *Laryngoscope*. 124, 2144-2150,

Jacobson, L. (2003) "Welcome to the virtual world." In Swadley, R. (Editor) On the cutting edge of technology. Carmel, Indianapolis.

Jaeger, M., Adair, D. (2012) "Construction safety simulations and students' perception of stress." 40th Annual SEFI Conference. Thessaloniki, Greece.

Jerome, C.J., Witmer, B.G. (2004) "Human performance in virtual environments: Effects of presence, immersive tendency, and simulator sickness." Proceedings of the Human Factors and Ergonomic Society (HFES) Conference, New Orleans.

Johns, C., Nuñez, D., Daya, M., Sellars, D., Casanueva, J., Blake, E. (2000) "The Interaction Between Individuals' Immersive Tendencies and the Sensation of Presence in a Virtual Environment." *Virtual Environments*. Proceedings of the Eurographics Workshop in Amsterdam. 65-74.

Johnson, A.M. (2011) “Charting a course for a successful Research Career. A Guide for Early Career Researchers.” 2nd Edition. Elsevier.

Johnson, R.B., Onwuegbuzie, A.J. (2004) “Mixed Methods Research: A Research Paradigm Whose Time Has Come.” *Educational Researcher*. 33: (70), 14-26.

Johnson, R.B., Onwuegbuzie, A.J., Turner, L.A. (2007) “Toward a Definition of Mixed Methods Research.” *Journal of Mixed Methods Research*. 1: (2), 112-133.

Jones, B., Kenward, M.G. (1989) “Design and Analysis of Cross-Over Trials.” 2nd Edition. Chapman and Hall/CRC. London.

Jordan, A., Carlile, O., Stack, A. (2008) “Approaches to Learning. A Guide For Teachers.” Open University Press. McGraw-Hill Education

Jordan, R.M. (1995) “The maturing years.” The Society and College of Radiographers, London.

Kahan, B.C., Morris, T.P. (2012) “Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis.” *British Medical Journal*. 345, e5840. [Downloaded March 2016]
http://www.bmj.com/content/345/bmj.e5840

Kak N, Burkhalter B, Cooper M. (2001) “Measuring the Competence of Healthcare Providers.” QA Operations Research Issue Paper. 2:1.

Kalaniti, K., Campbell, D.M. (2015) “Simulation-based Medical Education: Time for a Pedagogical Shift.” *Indian Pediatrics.* 52: (1), 41-45.

Källström, R. (2010) “Construction, Validation and Application of a Virtual Reality Simulator for the Training of Transurethral Resection of the Prostate.” Linköping University medical dissertations. No. 1167.

Kaplan, R. M., Bush, J. W., Berry, C.C. (1976) “Health Status: Types of Validity and the Index of Well-being.” *Health Services Research.* 11: (4), 478-507.

Kenney, P.A., Wszolek, M.F., Gould, J.J., Libertino, J.A., Molnzadeh, A. (2009) “Face, Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery.” *Urology*. 73: (6), 1288-1292.

Khan, F.M., Deibel, F.C., Soleimani-Meigooni, A. (1985) “Obliquity correction for electron beams.” *Medical Physics.* 12: (6), 749-753.

Khan, F.M., Gibbons, J.P. (2014) “Khan's; The Physics of Radiation Therapy.” 5th Edition. Lippincott Williams and Wilkins.

Khemani, S., Arora, A., Singh, A., Tolley, N., Darzi, A. (2012) “Objective skills assessment and construct validation of a virtual reality temporal bone simulator.” *Ontology & Neurotology*. 7: 1225-31.

Kimberlin, C.L., Winterstein, A.G. (2008) “Validity and reliability of measurement instruments used in research.” *American Journal of Health-System Pharmacy*. 65: 2276-84.

Kincheloe, J.L., Steinberg, S.R. (1993). “A tentative description of post-formal thinking: The critical confrontation with cognitive theory.” In Steinberg, S.R., Kincheloe, J.L., Hinchey, P.H. (Editors) The post-formal reader: Cognition and Education. New York: Falmer Press.

Kirschner, P., Strijbos, J., Kreijns, K., & Beers, P. J. (2004). “Designing electronic collaborative learning environments.” *Educational Technology Research and Development*. 52: (3), 47-66.

Klassen, S. (2006) “Contextual Assessment in Science Education: Background, Issues, and Policy.” *Science Education*. 90: (5), 820-851.

Kline, P. (1993) “Handbook of Psychological Testing.” Routledge.

Kolozsvari, N.O., Feldman, L.S., Vassiliou, M.C., Demyttenaere, S., Hoover, M.L. (2011) “Sim One, Do One, Teach One: Considerations in Designing Training Curriculua for Surgical Simulation.” *Journal of Surgical Education*. 68: 421-427.

Kondracki, N.L., Wellman, N.S., Amundson, D.R. (2002) “Content Analysis: Review of Methods and Their Applications in Nutrition Education.” *Journal of Nutrition Education and Behavior*. 34: 224-230.

Kozhevnikov, M. (2007) “Cognitive Styles in the Context of Modern Psychology: Toward an Integrated Framework of Cognitive Style.” *Psychological Bulletin*. 133: (3), 464-481.

Kroutter Jr, P.J. (2010) “The influence of field dependence /independence, gender, and experience on navigational behavior and configurational knowledge acquisition in a desktop virtual reality environment.” PhD thesis. Oklahoma State University.

Krueger, R.A. (1994) “Focus groups: A practical guide for applied research.” 2nd Edition. Thousand Oaks, CA: Sage.

Kunkler, T. (2003) in Bomford, C. K., Kunkler, I. H. “Walter and Miller’s Textbook of Radiotherapy.” 6th Edition. Churchill Livingstone.

Lainema, T. (2003) “Implications of Constructivism, for Computer-based Learning. Proceedings of European conference of Information Systems.” Neapel, Italy.

Larsen, C.R., Grantcharov, T., Aggarwal, R., Tully, A., Sørensen, L.J., Dalsgaard, T., Ottesen, B. (2006) “Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator.” *Surgical Endoscopy*. 20: 1460-1466.

Lateef, F. (2010) “Simulation-based learning: Just like the real thing.” *Journal of Emergencies, Trauma and Shock*. 3: (4), 348-352.

LeBlanc, V.R. (2009) "The Effects of Acute Stress on Performance: Implications for Health Professions Education." *Academic Medicine*. 84: (10 Supplement) S25-S33.

Lederman, M. (1981) "The early history of Radiotherapy: 1895-1939." *International Journal of Radiation Oncology, Biology and Physics*. 7: (5), 639-48.

Leggett, T. (2015) "Competency-Based Education: A brief overview." Radiologic Technology. 86: (4), 445-448.

Levin, S.L., Mohamed, F.B., Platek, S.M. (2005) "Common ground for spatial cognition? A behavioral and fMRI study of sex differences in mental rotation and spatial working memory." *Evolutionary Psychology*. 3: 227-254.

Ling, Y., Nefs, H.T., Brinkman, W.P., Qu, C., Heynderickx, I. (2013) "The relationship between individual characteristics and experienced presence." *Computers in Human Behavior*. 29: (4), 1519-1530.

Liu, D., Macchiarella, N.D., Vincenzi, D.A. (2009) in Vincenzi, D.A., Wise, J.A. Mouloua, M., Hancock, P.A. "Human factors in simulation and training." CRC Press.

Lombard, M., Ditton, T. (1997) "At the Heart of It All: The Concept of Presence." *Journal of Computer-Mediated Communication*. 3: (2). [Downloaded October 2014] http://jcmc.indiana.edu/vol3/issue2/lombard.html

Looney J.W. (2011) "Integrating Formative and Summative Assessment. Progress toward a seamless system?" OECD Education Working Papers, No. 58, OECD Publishing.

Loukas, C., Nikiteas, N., Kanakis, M., Georgiou, E. (2011) "Deconstructing laparoscopic competence in a virtual reality simulation environment." *Surgery*. 149: (6), 750-60.

Ma, I.W.Y., Brindle, M.E., Ronksley, P.E., Lorenzetti,D.L., Ghali, W.A.G. (2011) "Use of simulation-based education to improve outcomes of central venous cathetization: A systematic review and meta-analysis." *Academic Medicine*. 86: (9), 1137-1147.

Ma, R., Kaber, D.B. (2006) “Presence, workload and performance effects of synthetic environment design factors.” *International Journal of Human-Computer Studies*. 64: 541-552.

Maagaard, M., Sorensen, J.L., Oestergaard, J., Dalsgaard, T., Grantcharov, T.P., Ottesen, B.S., Larsen, C.R. (2011) “Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer.” *Surgical Endoscopy*. 25: 722-727.

Mackay, S. (2003) “Problem-based learning in radiographer education: testing the water before taking the plunge.” *Critical Encounters: Scholarly Approaches to Learning & Teaching.* Continuing Professional Development Series 7. York: Higher Education Academy 93-108

Maclure, M., Mittleman, M.A. (2000) “Should we use a case-Crossover Design?” *Annual Review of Public Health*. 21: 193-221.

Madden, C.A., Mitchell, V. A. (1993) “Professions, Standards and Competence: a survey of continuing education for the professions.” Bristol: University of Bristol.

Maddux, C., Johnson, D., Willis, J. (2001). “Educational computing: Learning with tomorrow’s technologies.” Needham Heights, MA: Allyn Bacon.

Maier, F.H., Grossler, A. (2000). “What are we talking about--A taxonomy of computer simulations to support learning.” *System Dynamics Review.* 6: (2), 135-148.

Mäntylä, T. (2013) “Gender Differences in Multitasking Reflect Spatial Ability.” *Psychological Science*. 24: (4), 514-520.

Maran, N.J., Glavin, R.J. (2003) “Low- to high-fidelity simulation - a continuum of medical education?” *Medical Education.* 37: (Suppl. 1): 22-28.

Markus, M.L., Silver, M.S. (2008) “A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit.” *Journal of the Association for Information Systems.* 9: (10/11), 609-632.

Martin, K.L., Hodgson, D. (2006) “The role of counselling and communication skills: how can they enhance a patient's ‘first day’ experience?” *Journal of Radiotherapy in Practice*. 5: (3), 157-164.

Marunić, G., Glažar, V. (2014) “Improvement and assessment of spatial ability in Engineering Education.” *Engineering Review*. 34: (2), 139-150.

Mathew, M. (2014) “Key factors in the promotion and obstruction of simulated learning in practice: an overview.” *Practice and Research in Education*. University of Bolton education research student journal. Issue 1.

Matthews, J.N.S., Henderson, R. (2013) “Two-period, two-treatment crossover designs subject to non-ignorable missing data.” *Biostatistics*. 14: (4), 626-638.

Mayring, P. (2000) “Qualitative Content Analysis.” Forum Qualitative Social Research FQS, 1: (2). [Downloaded November 2013] http://www.qualitative-research.net/fqs-texte/3-07/07-3-26-e.htm

McCarthy, J.P., Anderson, L. (2000) “Active Learning Techniques Versus Traditional Teaching Styles: Two Experiments from History and Political Science.” *Innovative Higher Education*. 24: (4), 279-294.

McComas, J., Pivik, J., Laflamme, M. (1998) “Current uses of virtual reality for children with disabilities.” *Studies in Health Technology and Informatics*. 58, 161-169.

McDougall, E.M., Corica, F.A., Boker, J.R., Sala, L.G., Stoliar, G., Borin, J.F., Chu, F.T., Clayman, R.V. (2006) “Construct Validity Testing of a Laparoscopic Surgical Simulator.” *Journal of the American College of Surgeons*. 202: (5), 779-787.

McGaghie, W.C, Issenberg S.B., Cohen, E.R., Barsuk, J.H., Wayne, D.B. (2011) “Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A met-analytical comparative review of the evidence.” *Academic Medicine*. 86: (6), 706-711.

McLellan, H. (2004) “Virtual Realities.” In Jonassen, D.H. (Editor) Handbook of Research on Educational Communications and Technology. 2nd Edition. Routledge.

McMahan, A. (2003) “Immersion, engagement and Presence. A Method for Analyzing 3D Videogames.” In Wolf, M.J.P., Perron, B. (Editors) The Video Game, Theory Reader. Routledge, Taylor & Francis Group.

McMillan, T.J. (2002) “Principles of radiobiology.” In Bomford, C.K., Kunkler, I.H. (Editors) Textbook of Radiotherapy. Churchill Livingstone. Edinburgh.

Meijer, F., Geudeke, B.L., van den Broek, E.L. (2009) “Navigating through Virtual Environments: Visual Realism Improves Spatial Cognition.” *Cyberpsychology and Behavior*. 12: (5), 517-521.

Mestre D, (2006) “Immersion et présence.” In Fuchs P, Moreau G (Editors) Traité de la Réalité Virtuelle. 3rd Edition. 1: 309-338. Presses de l’Ecole des Mines, Paris. Translation [Downloaded December 2014]
http://www.ism.univmed.fr/mestre/pub/Pres_2005.pdf.

Michelson, J. D., Manning, L. (2008) “Competency assessment in simulation-based procedural education.” *The American Journal of Surgery*. 196: 609-615.

Middlemiss, J. (1973) “Radiological education of the future.” *British Journal of Radiology*. 46: (550), 803-807.

Milgram, P., Takemura, H., Utsumi, A., Kishino, F. (1994) “Augmented Reality: A class of displays on the reality-virtuality continuum.” Proceedings of SPIE 2351. Telemanipulator and Telepresence Technologies. 282-292.

Miller, D.I., Halpern, D.F. (2014) “The new science of cognitive sex differences.” *Trends In Cognitive Sciences*. 18: (1), 37-45.

Min Lee, K. (2004) “Presence, Explicated.” *Communication Theory*. 14: (1), 27-50.

Min-Tae, K., Hae-Kag, L., Yeong-Cheol, H., Jae-Hwan, C. (2014) "A Study on Effective Source-Skin Distance using Phantom in Electron Beam Therapy." *Journal of Magnetics*. 19: (10), 15-19.

Mistry, M., Parkin, D.M., Ahmed, A.S., Sasieni, P. (2011) "Cancer incidence in the United Kingdom: projections to the year 2030." *British Journal of Cancer*. 105: (11), 1795-1803.

Mitchell, G. (2013) "The Rationale for Fractionation in Radiotherapy." *Clinical Journal of Oncology Nursing*. 17: (4), 412-417.

Moore, P. (1995) "Learning and teaching in virtual worlds: Implications of virtual reality for education." *Australian Journal of Educational Technology*. 11: (2), 91-102.

Morgan, P.J., Cleave-Hogg, D., Guest, C.B. (2001) **"**A Comparison of Global Ratings and Checklist Scores from an Undergraduate Assessment Using an Anesthesia Simulator." *Academic Medicine*. 76: (10), 1053-1055.

Moroney, W.F., Lilienthal, M.G. (2008) "Human factors in Simulation and Training: An Overview." In Hancock, P.A., Vincenzi, D.A, Wise, J.A., Mouloua, M. (Editors) Human Factors in Simulation and Training. CRC Press.

Mould, R. F. (2011) "X-rays in 1896-1897." *Journal of Oncology*. 61: (6), 100e-109.

Muncey, T. (2009) "An overview of mixed methods research." *Journal of Research in Nursing*. 14: (2), 187-188.

Mundell, W.C., Kennedy, C.C., Szostek, J.H., Cook, D.A. (2013) "Simulation technology for resuscitation training: A systematic review and meta-analysis." *Resuscitation*. 84, 1174-1183.

Murray, C.D., Fox, J. and Pettifer, S. (2007) "Absorption, dissociation, locus of control and presence in virtual reality." *Computers in Human Behavior*, 23: (3), 1347-1354.

Myers, R.D. (2012) "Analyzing interaction patterns to verify a simulation/game model." PhD dissertation. School of Education, Indiana University.

Nanjappa, A., Grant, M.M. (2003) "Constructing on Constructivism: The Role of Technology." *Electronic Journal for the Integration of Technology in Education*.

National Radiotherapy Advisory Group (NRAG) Report (2007) "Radiotherapy: developing a world class service for England." Department of Health.

Nestel D., Tabak, D., Tierney, T., Layat-Burn, C., Robb, A., Clark, S., Morrison, T., Jones, N., Ellis, R., Smith, C., McNaughton, N., Knickle, K., Higham, J., Kneebone, R. (2011) "Key challenges in simulated patient programs: an international comparative case study." *BMC Medical Education*. 11: (69), 1-10.

Newman, D.A. (2014) "Missing Data: Five practical guidelines." *Organizational Research Methods*. 17: (4), 372-411.

Nieder, C., Baumann, B. (2011) "Fractionation Concepts." In Nieder, C., Langendijk, J.A. (Editors) Re-Irradiation: New Frontiers, Medical Radiology. Radiation Oncology. Springer-Verlag.

Nixon, S. (2001) "Professionalism in Radiography." *Radiography*. 7: (1), 31-35.

Noeth, R.J., Volkov, B.B. (2004) "Evaluating the Effectiveness of Technology in Our Schools. ACT Policy Report." [Downloaded July 2014] http://www.act.org/research/policymakers/pdf/school_tech.pdf

Northwest Radiation Oncology (2013) "What are the steps involved in radiation treatment?" [Downloaded July 2014] http://www.radiationdocs.org/what-are-the-steps-involved-in-radiation-treatment.html

Nowak, K., Biocca, F. (2003) "The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments." *Presence*. 12: (5), 481–494.

Office for National Statistics (2012) "Cancer Incidence and Mortality in the United Kingdom, 2008-10." Crown copyright.

OFT (2014) "Universities' Terms and Conditions. An OFT report." OFT.

Ogden, P.E., Cobbs, L.S., Howell, M.R., Sibbitt, S.J., DiPette, D.J. (2007) "Clinical simulation: importance to the internal medicine educational mission." *American Journal of Medicine*. 120: 820-824.

Okuda. Y., Bryson, E.O., DeMaria, S. Jr, Jacobson, L., Quinones, J., Shen, B., Levine, A.I. (2009) "The utility of simulation in medical education: what is the evidence?" *Mount Sinai Journal of Medicine.* 76: 330-343.

Oluwatayo, J. A. (2012) "Validity and Reliability Issues in Educational Research." *Journal of Educational and Social Research.* 2: (2), 391-400.

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi, S., David, D. (2012) "Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis." *Depression and Anxiety*. 29: (2), 85-93.

Orb, A., Eisenhauer, L., Wynaden, D. (2001) "Ethics in qualitative research." Journal of Nursing Scholarship. 33: (1), 93-96.

Orton, C.G. (2013) "Radiation Dosimetry Physical and Biological Aspects." Springer.

Oswald, D., Sherratt, F., Smith, S. (2014) "Handling the Hawthorne effect: The challenges surrounding a participant observer." *Review of Social Studies*. 1: (1), 53-73.

Owen, H. (2012) "Early use of simulation in medical education." *Simulation in Healthcare*. 7: (2), 102-116.

Owen, H., Mugford, B., Follows, V., Plummer, J.L. (2006) “Comparison of three simulation-based training methods for management of medical emergencies.” *Resuscitation*. 71: 204-211.

Palfreyman, D. (2013) “Quality & consumerism in higher Education.” *Perspectives: Policy and Practice in Higher Education*. 17: (3), 107-111.

Pantelidis, V. S. (2010) “Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality.” *Themes in Science and Technology Education*. 3: (1-2), 59-70.

Pasveer, B. (1989) “Knowledge of shadows: the introduction of X-ray images in medicine.” *Sociology of Health and Illness*. 11: (4), 360-381.

Patel, A.D., Gallagher, A.G., Nicholson, W.J., Cates, C.U. (2006) “Learning Curves and Reliability Measures for Virtual Reality Simulation in the Performance Assessment of Carotid Angiography.” *Journal of the American College of Cardiology*. 47: (9), 1796-1802.

Pedhazur, E. J., Schmelking, L. (1991) “Measurement, Design, and Analysis: An Integrated Approach.” Student Edition. Lawrence Erlbaum Associates.

Peña, J., Blackburn, K. (2013) “The Priming Effects of Virtual Environments on Interpersonal Perceptions and Behaviors.” *Journal of Communication*. 63, 703–720.

Perkins, G.D. (2007) “Simulation in resuscitation training.” *Resuscitation*. 73, 202-211.

Persky, S., Kaphingst, K.A., McCall, C., Lachance, C., Beall, A.C., Blascovich, J. (2009) “Presence Relates to Distinct Outcomes in Two Virtual Environments Employing Different Learning Modalities.” *Cyberpsychology & Behavior*. 12: (3), 263-268.

Pictor, A. (2003) “Analysing Binary Outcome Data from a Crossover Design Study using the SAS® System.” SAS Views.
[Downloaded January 2014] http://www.lexjansen.com/views/2003/statpharm/st02.pdf

Pietsch, S., Jansen, P. (2012) “Different mental rotation performance in students of music, sport and education.” *Learning and Individual Differences*. 22: (1), 159-163.

Pilcher, J.W., Bedford, L. (2011) “Willingness and Preferences of Nurses Related to Learning With Technology.” *Journal for Nurses in Professional Development*. 27: (3), E10-E16.

Polit, D.F., Tatano Beck, C. (2012) “Nursing Research. Generating and Assessing Evidence for Nursing Practice. 9th Edition.” Lippincott, Williams and Wilkins, Philadelphia.

Porter-O’Grady, T. (2001). “Profound change: 21st century nursing.” *Nursing Outlook*. 49: (4), 184.

Potgieter, T., Van Der Merwe, R.P. (2002) “Assessment in the workplace: A competency-based approach.” *South African Journal of Industrial Psychology*. 28: (1), 60-66.

Privitera, G.J. (2011) “Statistics for the Behavioral Sciences.” Sage publications.

Proulx, J. (2006) “Constructivism: A re-equilibration and clarification of the concepts, and some potential implications for teaching and pedagogy.” *Radical Pedagogy*. 8: (1), [Downloaded August 2014]
http://radicalpedagogy.icaap.org/content/issue8_1/proulx.html

Prystowsky, J.B., Regehr, G., Rogers, D.A., Loan, J.P., Hiemenz, L.L., Smith, K.M. (1999) “A Virtual Reality Module for Intravenous Catheter Placement.” American Journal of Surgery. 177: 171-175.

Rabiee, F. (2004) “Focus-group interview and data analysis.” Proceedings of the Nutrition Society (2004), 63: 655-660.

Raffe, D. (1992) “Modularisation in Initial Vocational Training: Recent developments in six European countries.” University of Edinburgh, Edinburgh, UK

Ramsden, P. (2003) “Learning to teach in Higher Education.” 2nd Edition. London. Routledge.

Ravassard, P., Kees, A., Willers, B., Ho, D., Aharoni, D., Cushman, J., Aghajan, Z.M., Mehta, M.R. (2013) “Multisensory Control of Hippocampal Spatiotemporal Selectivity.” Science. 340, 1342-1346. [Downloaded January 2015] http://www.sciencemag.org/

Rideout, V.J., Foehr, U.G., Roberts, D.F. (2010) “Generation M^2: Media in the Lives of 8 to 18 Year Olds.” Kaiser Family Foundation.

Riva, G., Mantovani, F. (2012) In Eichenberg, E. Editor “Virtual Reality in Psychological, Medical and Pedagogical Applications.” Intech. [Downloaded March 2016]
http://www.intechopen.com/books/virtual-reality-in-psychological-medical-and-pedagogical-applications/being-there-understanding-the-feeling-of-presence-in-a-synthetic-environment-and-its-potential-for-c

Riva, G., Mantovani, F. (2014) in Riva, G., Waterworth, J., Murray, D. (2014) “Interacting with Presence HCI and the Sense of Presence in Computer-mediated Environments.” De Gruyter Open. Warsaw.

Roach, K.E. (2006) “Measurement of Health Outcomes: Reliability, Validity and Responsiveness.” *Journal of Prosthetics and Orthotics*. 18: (1S), 8-12.

Roberts, D. F., Foehr, U. G., Rideout V. J. (2005) “Generation M: Media in the lives of 8–18 year-olds.” Kaiser Family Foundation.

Romano, D.M., Brna, P. (2001). “Presence and reflection in training: Support for learning to improve quality decision-making skills under time limitations.” *CyberPsychology & Behavior*, 4: (2), 265-277.

Rosenfield, R.M. (2010) "How to review journal manuscripts." *Otolaryngology–Head and Neck Surgery*. 142, 472-486.

Royal College of Radiologists, Society and College of Radiographers, Institute of Physics and Engineering in Medicine. (2008) "On target: Ensuring geometric accuracy in Radiotherapy." The Royal College of Radiologists.

Ryan, C.A., Walshe, N., Gaffney, R., Shanks, A., Burgoyne, L., Wiskin, C.M. (2010) "Using standardized patients to assess communication skills in medical and nursing Students." *BMC Medical Education*. 10: (24), 1-8.

Saenz, M.J., Cano, J.L. (2009) "Experiential learning through simulation games: An empirical study." *International Journal of Engineering Education*. 25: (2), 296-307.

Safaei and Shafieiyoun, (2013) "Enhancing Learning within the 3-D Virtual Learning Environment." *Journal of Knowledge Management, Economics and Information Technology*. 3: (3), 1-18.

Saldaňa, J. (2012) "The Coding Manual for Qualitative Researchers." SAGE Publications Ltd; 2nd Edition.

Salkind, N.J. (2010) "Encyclopedia of Research Design." [Downloaded October 2014] http://0-dx.doi.org.wam.city.ac.uk/10.4135/9781412961288

Sas, C., O'Hare, G.M.P. (2003) "Presence equation: An investigation into cognitive factors underlying presence." *Presence: Teleoperators and Virtual Environments*. 12: (5), 523-537.

Scalese, R.J., Obeso, V.T., Issenberg, S.B. (2008) "Simulation Technology for Skills Training and Competency. Assessment in Medical Education." *Journal of General Internal Medicine*. 23 (Suppl 1): 46-9.

Schreuder, H.W.R., Persson, J.E.U., Wolswijk, R.G.H., Ihse, I., Schijven, M.P., Verheijen, R.H.M. (2014) “Validation of a Novel Virtual Reality Simulator for Robotic Surgery.” *The Scientific World Journal*. [Downloaded December 2014] http://dx.doi.org/10.1155/2014/507076

Schreuder, H.W.R., van Dongen, K.W., Roeleveld, S.J., Schijven, M.P., Breders, A.M.J. (2009) “Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery.” *American Journal of Obstetric Gynecology*. 200: (540), e1-e8.

Schuemi, M.J., van der Mast, C.A.P.G., Abel, B., Emmelkamp, P.M.G., Krijn, M. (2005) “The effect of locomotion technique on presence and fear and usability in a virtual environment.” Proc.Euromedia. Al-Akaidi, M. & Rothkrantz, L. (Eds.) Toulouse Apil11-13. 129-135.

Schute, V.J. (2007) “Focus on Formative Feedback.” Educational Testing Service Research Report.

Schuwirth, L.W.T. van der Vleuten C.P.M. (2003) “The use of clinical simulations in assessment.” *Medical Education*. 37: (Suppl. 1), 65-71.

Schuwirth, L.W.T., van der Vleuten, C.P.M. (2011) “Programmatic assessment: From assessment of learning to assessment for learning.” *Medical Teacher*. 33, 478-485.

Senn, S.S. (2002) “Cross-over Trials in Clinical Research (Statistics in Practice).” 2nd Edition. Wiley-Blackwell.

Sherman, W.R., Craig, A.B. (2003) “Understanding Virtual Reality: Interface, Application, and Design.” Morgan Kaufmann.

Shetty, S., Panait, L., Baranoski, J., Dudrick, S.D., Bell, R.L., Roberts, K.E., Duffy, A.J. (2012) “Construct and face validity of a virtual reality-based camera navigation curriculum.” *Journal of Surgical Research*. 177: (2) 191-195.

Siebeck, M., Schwald, B., Frey, C., Roding, S., Stegmann, K., Fischer, F. (2011) "Teaching the rectal examination with simulations: effects on knowledge acquisition and inhibition." *Medical Education*. 45: (10), 1025-31.

Sim J. and Radloff, A. (2009) "Profession and professionalization in medical radiation science as an emergent profession." *Radiography*. 15: (3), 203-208.

Simpson, P.M., Hamer, R.M., Lensing, S.L. (1999) "Cross crossover studies off your list." Paper 221. 24th, SAS Users Group International Conference Paper. 1303-1309.

Siriaraya, P., Ang, C.S. (2012) "Age differences in the perception of social presence in the use of 3D virtual world for social interaction." *Interacting with Computers*. 24: (4), 280-291.

Slater, M. (1999) "Measuring presence: A response to the Witmer and Singer presence questionnaire." *Presence-Teleoperators and Virtual Environments*. 8: (5), 560-565.

Slater, M., Wilbur S. (1997) "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments." *Presence: Teleoperators and Virtual Environments*. 6: (6), 603-616.

Slevin, M.L., Nichols, S.E, Downer, S.M., Wilson, P., Lister, T.A., Arnott, S., Maher, J., Souhami, R.L., Tobias, J.S., Goldstone, A.H., Cody, M. (1996) "Emotional support for cancer patients: what do patients really want?" *British Journal of Cancer*. 74: (8), 1275-1279.

Smeby, J.C. (2012) "How Can Qualitative and Quantitative Data Sets Be Linked?" Norwegian Educational Research towards 2020 - UTDANNING2020.

Sokolowski, J.A. (2011) "The practice of Modelling and Simulation: Tools of the Trade." In Sokolowski, J.A., Banks, C.M. (Editors) "Modelling and Simulation in the Medical and Health Sciences." Wiley-Blackwell.

SOR (2003) Society of Radiographers. Identity Statement. [Downloaded June 2014] http://www.aim25.ac.uk/cgi-bin/vcdf/detail?coll_id=7276&inst_id=100

Spence, I., Yu, J.J., Feng, J., Marshman, J. (2009) "Women Match Men When Learning a Spatial Skill." *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 35: (4), 1097-1103.

Stanton, D., Wilson, P., Foreman, N. (1996) "Using virtual reality environments to aid spatial awareness in disabled children." *European Conference on Disability, Virtual Reality and Associate Technology*. Maidenhead. UK.

Steel, G.G., McMillan, T.J., Peacock, J.H. (1989) "The 5Rs of radiobiology." *International Journal of Radiation Biology*. 56: 1045-1048.

Steinberg, A.D., Bashook, P.G., Drummond, J., Ashrafi, S., Zefran, M. (2007) "Assessment of Faculty Perception of Content Validity of PerioSim©, a Haptic-3D Virtual Reality Dental Training Simulator." Journal of Dental Educucation. 71: (12), 1574-82.

Stewart, D.W., Shamdasani, P.N., Rook, D.W. (2007) "Focus Groups. Theory and Practice." Thousand Oaks: Sage Publications. OpenURL

Still, J.D. (2009) "Conceptualizing design affordances from a cognitive perspective." Graduate Theses and Dissertations. Paper 10533.

Strangman, N., Hall, T., Meyer, A. (2003) "Virtual Reality/Computer Simulations and the Implications for UDL Implementation." NCAC Curriculum Enhancement Report.

Strydom, W., Parker, W., Olivares, M. (2005) "Electron beams: Physical and Clinical Aspects." In Podgorsak, E.B. (Editor) "Radiation Oncology Physics: A Handbook for Teachers and Students." International Atomic Energy Agency. Vienna.

Stunt, J.J., Wulms, P.H., Kerkhoffs, G.M., Dankelman, J., van Dijk, C.N., Tuithoh, G.J.M. (2014) “How valid are commercially available medical simulators?” *Advances in Medical Education and Practice*. 5: 385-395.

Suit, H D., Loeffler, J. S. (2011) Radiation Oncology and the MGH 1896–1945. Springer.

Sun, H.M., Cheng, W.L. (2009) “The Input-Interface of Webcam Applied in 3D Virtual Reality Systems.” *Computers and Education*. 53: 4, 1231-1240.

Suntharalingam, N., Podgorsak, E.B., Tőlli, H. (2005) “Brachytherapy: physical and clinical aspects.” In Podgorsak, E.B. (Editor) “Radiation Oncology Physics: A Handbook for Teachers and Students.” International Atomic Energy Agency. Vienna.

Suresh, K.P. (2011) “An overview of randomization techniques: An unbiased assessment of outcome in clinical research.” *Journal of Human Reproductive Sciences*. 4: (1), 8-11.

Swan, K. (2005) “A Constructivist model for thinking about learning online.” In Bourne, J., Moore, J.C. (Editors) Elements of Quality Online Education: Engaging Communities. Needham, MA: Sloan-C.

Ternier, S., Klemke, R., Kalz, M., Van Ulzen, P., Specht, M. (2012). “ARLearn: augmented reality meets augmented virtuality [Special issue].” *Journal of Universal Computer Science – Technology for Learning Across Physical and Virtual Spaces*. 18: (15), 2143-2164.

Thijssen, A.S., Schijven, M.P. (2010) “Contemporary virtual reality laparoscopy simulators quicksand or solid grounds for assessing surgical trainees?” *The American Journal of Surgery*. 199, 529-541.

Thomas, D.R. (2006) “A General Inductive Approach for Analyzing Qualitative Evaluation Data.” *American Journal of Evaluation*. 27: (2), 237-246.

Thompson, C., Len Dalgleish, L., Bucknall, T., Estabrooks, C., Hutchinson, A.M., Fraser, K., de Vos, R., Binnekade, J., Barrett, G., Saunders, J. (2008) "The Effects of Time Pressure and Experience on Nurses' Risk Assessment Decisions. A Signal Detection Analysis" *Nursing Research*. 57: (5), 302-311.

Thompson, P.W. (2000) "Radical Constructivism: Reflections and Directions." In Steffe, L.P., Thompson, P.W. (Editors) Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld. Falmer Press.

Threlfall, D.K. (1999) "Using focus groups as a consumer research tool." *Journal of Marketing Practice: Applied Marketing Science*. 5: (4), 102-105.

Thurman, R., Mattoon, J.S. (1994) "Virtual reality: Toward fundamental improvements in simulation-based training." *Educational Technology*. 34: (8), 56-64.

Thwaites, D., McKenzie, A. (2007) "Electron Beams." in Mayles, O., Nahum, A., Rosenwald, J.C. (Editors) Handbook of radiotherapy Physics. Theory and Practice. Taylor and Francis.

TLRP (2010) "Effective learning and teaching in UK higher education: A commentary by the Teaching and Learning Research Programme." London. TLRP. [Downloaded September 2014] http://www.tlrp.org/pub/documents/UKHEfinal.pdf

Tortell, R., Luigi, D.P., Morie, J., Dozois, A. (undated) "Effects of priming and videogame play habits on responses to a virtual environment." Ebook [Downloaded September 2014] http://projects.ict.usc.edu/see/publications/Priming_Soldier.pdf

Tsai, W.W., Fung, C.P., Tsai, S.L., Jeng, M.C., Doong, J.L. (2008) "The assessment of stability and reliability and validity of a virtual reality-based intravenous injection simulator." Computers, Informatics, Nursing. 26: (4), 221-226.

Tuli, F. (2010) "The Basis of Distinction Between Qualitative and Quantitative Research in Social Science: Reflection on Ontological, Epistemological and Methodological Perspectives." Ethiopian Journal of Education and Sciences, 6: (1), 97-108

Usoh, M., Catena, E., Arman, S., Slater, M. (2000) "Using Presence Questionnaires in Reality." *Presence: Teleoperators and Virtual Environments*. 9: (5), 497-503.

Van Baren, J., Ijsselsteijn, W. (2004) "Measuring Presence: A Guide to Current Measurement Approaches." OmniPres Project IST-2001-39237 [Downloaded October 2014] http://www.informatik.umu.se/~jwworth/PresenceMeasurement.pdf

Van Herzeele, I., Aggarwal, R., Choong, A., Brightwell, R., Vermassen, F.E., Cheshire, N.J. (2007) "Virtual reality simulation objectively differentiates level of carotid stent experience in experienced interventionalists." *Journal of Vascular Surgery*. 46: 855-63.

Van Schaik, P., Turnbull, T., Van Wersch, A., Drummond, S. (2004) "Presence within a mixed reality environment." *CyberPsychology & Behavior*, 7: (5), 540-552.

Vandewater, D.A. (2004) "Best Practices in Competence Assessment of Health Professionals." Background Policy Paper. College of Registered Nurses of Nova Scotia.

Ventura, M., Shute, V., Wright, T., Zhao, W. (2013) "An investigation of the validity of the virtual spatial navigation assessment." Frontiers in Psychology. 4: 852. [Downloaded October 2014] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861692/

Verhagen, T. Feldberg, F., van den Hooff, B., Meents, S., Merikivi, J. (2012). "Understanding users' motivations to engage in virtual worlds: A multipurpose model and empirical testing." *Computers in Human Behavior*. 28: 484-495

Viera, A. J., Garrett, J. M. (2005) "Understanding Interobserver Agreement: The Kappa Statistic." *Family Medicine.* 37: (5), 360-3.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). "Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables." *Psychological Bulletin*. 117, 250-270.

Walden, G.R. (2012) "Focus Group Research." Sage Publications Ltd.

Walker, W. (2005) “The strengths and weaknesses of research designs involving quantitative measures.” *Journal of Research in Nursing*. 10: (5), 571-582.

Wallston, K.A. (2005) “The Validity of the Multidimensional Health Locus of Control Scales.” *Journal of Health Psychology*. 10: (4), 623-631.

Walsh, L.J., Chai, L., Farah, C.S., Ngo, H., Eves, G. (2010) “Use of Simulated Learning Environments in Dentistry and Oral Health Curricula.” Report for Health Workforce Australia. [Downloaded September 2014] www.hwa.gov.au/work…/simulated-learning-environments-sles.

Wang, W. Tan, A.H., Teow, L.N., Tan, Y.S. “Declarative-Procedural Memory Interaction. in Learning Agentsin Lomuscio, A., Scerri, P., Bazzan, A., Huhns, M. (Editors) Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France. International Foundation for Autonomous Agents and Multiagent Systems.

Washington, C.M., Leaver, D.T. (2009) “Principles and Practice of Radiation Therapy.” 3rd Edition. Mosby.

Wellek, S., Blettner, M. (2012) “On the Proper Use of the Crossover Design in Clinical Trials: Part 18 of a Series on Evaluation of Scientific Publications.” *Deutsches Ärzteblatt International*. 109: (15), 276-81.

Westlake, P. (2011) “The Experiential Learning Model.” [Downloaded February 2015] http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.balticuniv.uu.se%2Findex.php%2Fcomponent%2Fdocman%2Fdoc_download%2F609-the-experiential-learning-model-&ei=0MflVNa0FdSP7AbY8IDYDw&usg=AFQjCNHxa_N3fpRbfEMeKsaUeeAp1N0nxA

Whitelock, D., Romano, D., Jelfs, A., Brna, P (2000) “Perfect presence: What does this mean for the design of virtual learning environments?” *Education and Information Technologies*. 5: (4), 277–289.

Winter, G. (2000). “A comparative discussion of the notion of validity in qualitative and quantitative research.” The Qualitative Report. 4 (3 & 4). [Downloaded September 2013] http://www.nova.edu/ssss/QR/QR4-3/winter.html

Wissmath, B., Weibel, D., Stricker, D. (2008) “When and How to Assess Subjective Overall Judgments of Presence?” PRESENCE 2008. Proceedings of the 11th Annual International Workshop on Presence, Padova, 16-18 October 2008.

Witmer, B.G., Jerome, C.J., Singer, M.J. (2005) “The Factor Structure of the Presence Questionnaire.” *Presence*. 14: (3), 298-312.

Witmer, B.G., Singer, M.J. (1998) “Measuring Presence in Virtual Environments: A Presence Questionnaire.” *Presence*. 7: (3), 225-240.

Witz, A. (2004) “Professions and Patriarchy.” Taylor and Francis.

Wolf, A. (2001) “Competence-Based Assessment.” From Raven, D., Stephenson, J., Lang, P. Competence in the learning society. Higher Education Authority. [Downloaded September 2014] http://www.heacademy.ac.uk/

Woods, A.J. (2005) “Compatibility of Display Products with Stereoscopic Display Methods.” International Display Manufacturing Conference. Taiwan. [Downloaded September 2014] www.cmst.curtin.edu.au

Wyatt, R.M., Jones, B.J., Dale, R.G. (2008) “Radiotherapy treatment delays and their influence on tumour control achieved by various fractionation schedules.” *British Journal of Radiology*. 81: (967), 549-63

Xiao, D., Jakimowicz, J.J., Albayrak, A., Buzink, S.N., Botden, S.M.B.I., Goossens, R.H.M. (2014) "Face, Content, and Construct Validity of a Novel Portable Ergonomic Simulator for Basic Laparoscopic Skills." *Journal of Surgical Education*. 71: (1), 65-72.

YouGov (2011) "A Review of 2011: Year of Radiotherapy." London.

Youngblut, C. (2003) "Experience of Presence in Virtual Environments." IDA Document D-2960 [Log H 03001849], 71 + 71.

Youngblut, C., Perrin, B.M. (2002). "Investigating the Relationship between Presence and Performance in Virtual Environments." *IMAGE 2002 Conference.* 8–12 July, Scottsdale, AZ.

Zahoric, P., Jenison, R.L. (1998). "Presence and being-in-the-world." *Presence: Teleoperators and Virtual Environments*. 7: (1), 78-89.

Zhang, J., Fu, X. (2015) "The Influence of Background Music of Video Games on Immersion." *Journal of Psychology & Psychotherapy*. 5: (4), 191. [Downloaded March 2016] http://www.omicsonline.org/open-access/the-influence-of-background-music-of-video-games-on-immersion-2161-0487-1000191.php?aid=58693

Appendix I – Ethics Committee Approval Letters

EXTERNAL AND STRATEGIC DEVELOPMENT SERVICES
uel.ac.uk/qa
Quality Assurance and Enhancement

UEL University of East London

Approved Documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Questionnaire	1.0	9 October 2012
Consent Form	1.0	9 October 2012
Risk Assessment	1.0	9 October 2012
Participant Information Sheet	1.0	9 October 2012

Approval is given on the understanding that the UEL Code of Good Practice in Research is adhered to.

With the Committee's best wishes for the success of this project.

Yours sincerely,

Merlin Harries
University Research Ethics Committee (UREC)
Quality Assurance and Enhancement
Telephone: 0208-223-2009
Email: researchethics@uel.ac.uk

Docklands Campus, University Way, London E16 2RD
Tel: +44 (0)20 8223 3322 Fax: +44 (0)20 8223 3394 MINICOM 020 8223 2853
Email: r.carter@uel.ac.uk

EXTERNAL AND STRATEGIC DEVELOPMENT SERVICES
uel.ac.uk/qa

Quality Assurance and Enhancement

20 October 2012

Dear David / John,

Project Title:	**Competency based assessment using virtual reality (VERT): Is it a realistic possibility?**
Researcher(s):	**David Flinton**
Principal Investigator:	**Professor John Preston**

I am writing to confirm that the application for the aforementioned proposed research study has now received ethical approval on behalf of University Research Ethics Committee (UREC).

Should any significant adverse events or considerable changes occur in connection with this research project that may consequently alter relevant ethical considerations, this must be reported immediately to UREC. Subsequent to such changes an Ethical Amendment Form should be completed and submitted to UREC.

Approved Research Site

I am pleased to confirm that the approval of the proposed research applies to the following research site.

Research Site	Principal Investigator / Local Collaborator
UEL Fieldwork	John Preston

Docklands Campus, University Way, London E16 2RD
Tel: +44 (0)20 8223 3322 Fax: +44 (0)20 8223 3394 MINICOM 020 8223 2853
Email: r.carter@uel.ac.uk

CITY UNIVERSITY LONDON

City Research Development & International Relations Office

Northampton Square
London, EC1V 0HB
United Kingdom
T +44 (0)20 7040 5060
F +44 (0)20 7040 3803
www.city.ac.uk

David Flinton
School of Health Sciences
City University London
Northampton Square
London
EC1V 0HB

24 October 2012

Dear Mr Flinton

Re: "Competency based assessment using virtual reality (VERT): Is it a realistic possibility?"

I am writing to you to confirm that you have been granted the permission to conduct research and recruit students at City University London, on the project detailed above. Senate Research Ethics Committee notes that the project has already been approved by the University Research Ethics Committee, at University of East London.

Should you have any further queries relating to this matter then please do not hesitate to contact me. On behalf of the Research Ethics Committee I do hope that the project meets with success and many thanks for your patience.

Kind regards

Anna Ramberg
Research Development Manager
Secretary to Research Ethics Committee

Email: Anna.Ramberg.1@city.ac.uk
Tel: 020 7040 3040

The University for business and the professions

Appendix II – Dissemination and Publications

1. 3rd VERT International user meeting. (2012) Oral Presentation. Kingston University.

2. International Conference on Virtual and Augmented Reality in Education. (2013) Poster presentation. Tenerife. Winner of the best poster prize.

   Paper linked to the above presentation is available on line, http://udv.ull.es/vare/data/vare2013_ID_62_poster%20PAPER.pdf

3. Health has got Talent. (2013) Oral Presentation. City University London.

4. 4th VERT International user meeting. (2014) Oral Presentation. University of Liverpool.

5. Network for Education in Healthcare Conference. (2015) Oral Presentation. Cambridge.

Appendix III – Letter of Invitation, Subject Information Sheet and Consent Letter

LETTER OF INVITATION/ RESEARCH PARTICIPANT INFORMATION SHEET

Competency based assessment using Virtual Reality.

Dear Student,
You are being asked to participate in a study that I am conducting as part of my doctoral thesis which is looking at the potential use of VERT as an assessment tool.

I am asking radiotherapy students to undertake two electron set-ups, one on VERT and one in a radiotherapy department. A small number of people will then be selected for a focus group.
Your participation would involve a maximum 3 sessions, 2 electron set-ups and a focus group meeting. The electron set-ups should each take about 15 minutes, the focus group no more than half an hour. There are also 3 short questionnaires you will need to complete.

Before you decide whether or not to participate, please read the Participation Information Sheet attached which provides further details about the study.

If you require more information, or wish to volunteer for this study, please either phone me on 0207 040 5688, or email me at: d.m.flinton@city.ac.uk

If on reading the Participant Information Sheet you decide you want to take part reply to this email to confirm this. If you do not wish to take part you need take no further action.

Yours Sincerely

D Flinton

David Flinton

Appendix IV – Survey Instruments

Subject No:

IMMERSIVE TENDENCIES QUESTIONNAIRE
(Witmer & Singer, Version 3.01)

Indicate your preferred answer by marking an "X" in the appropriate box of the seven point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply. For example, if your response is once or twice, the second box from the left should be marked. If your response is many times but not extremely often, then the sixth (or second box from the right) should be marked.

1. Do you easily become deeply involved in movies or TV dramas?

NEVER			OCCASSIONALLY			OFTEN

2. Do you ever become so involved in a television program or book that people have problems getting your attention?

NEVER			OCCASSIONALLY			OFTEN

3. How mentally alert do you feel at the present time?

NOT ALERT			MODERATLELY			FULLY ALERT

4. Do you ever become so involved in a movie that you are not aware of things happening around you?

NEVER			OCCASSIONALLY			OFTEN

5. How frequently do you find yourself closely identifying with the characters in a story line?

NEVER			OCCASSIONALLY			OFTEN

6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?

NEVER			OCCASSIONALLY			OFTEN

8. How physically fit do you feel today?

NOT FIT			MODERATELY FIT			EXTREMELY FIT

9. How good are you at blocking out external distractions when you are involved in something?

NOT VERY GOOD			SOMEWHAT GOOD			VERY GOOD

10. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

NEVER			OCCASSIONALLY			OFTEN

11. Do you ever become so involved in a daydream that you are not aware of things happening around you?

NEVER OCCASSIONALLY OFTEN

12. Do you ever have dreams that are so real that you feel disoriented when you awake?

NEVER OCCASSIONALLY OFTEN

13. When playing sports, do you become so involved in the game that you lose track of time?

NEVER OCCASSIONALLY OFTEN

14. How well do you concentrate on enjoyable activities?

NOT AT ALL MODERATELY WELL VERY WELL

15. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

NEVER OCCASSIONALLY OFTEN

16. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

NEVER OCCASSIONALLY OFTEN

17. Have you ever gotten scared by something happening on a TV show or in a movie?

NEVER OCCASSIONALLY OFTEN

18. Have you ever remained apprehensive or fearful long after watching a scary movie?

NEVER OCCASSIONALLY OFTEN

19. Do you ever become so involved in doing something that you lose all track of time?

NEVER OCCASSIONALLY OFTEN

Finally some demographic questions

Your age ______________

Your gender Male ☐ Female ☐

Your year 1st ☐ 2nd ☐ 3rd ☐

Thank You

Subject No:

PRESENCE QUESTIONNAIRE

(Witmer& Singer, Vs. 4.0)

Characterize your experience in the environment, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

What system setting did you use?

3D	☐
3D with tracking	☐
LINAC	☐

1. How much were you able to control events?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Somewhat — Completely

2. How responsive was the environment to actions that you initiated (or performed)?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not Responsive — Moderately Responsive — Completely Responsive

3. How natural did your interactions with the environment seem?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Extremely Artificial — Borderline — Completely Natural

4. How much did the visual aspects of the environment involve you?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Somewhat Completely

5. How much did the auditory aspects of the environment involve you?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Somewhat Completely

6. How natural was the mechanism which controlled movement through the environment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Extremely Artificial Borderline Completely Natural

7. How compelling was your sense of objects moving through space?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Moderately Compelling Very Compelling

8. How much did your experiences in the virtual environment seem consistent with your real-world experiences?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not Consistent — Moderately Consistent — Very Consistent

9. Were you able to anticipate what would happen next in response to the actions that you performed?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Somewhat — Completely

10. How completely were you able to actively survey or search the environment using vision?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Somewhat — Completely

11. How well could you identify sounds?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Somewhat — Completely

12. How well could you localize sounds?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Somewhat — Completely

13. How well could you actively survey or search the virtual environment using touch?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Somewhat Completely

14. How compelling was your sense of moving around inside the virtual environment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not Compelling Moderately Compelling Very Compelling

15. How closely were you able to examine objects?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Pretty Closely Very Closely

16. How well could you examine objects from multiple viewpoints?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Somewhat Extensively

17. How well could you move or manipulate objects in the virtual environment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Somewhat Extensively

18. How involved were you in the virtual environment experience?

☐	☐	☐	☐	☐	☐	☐
Not Involved			Mildly Involved			Completely Engrossed

19. How much delay did you experience between your actions and expected outcomes?

☐	☐	☐	☐	☐	☐	☐
No Delays			Moderate Delays			Long Delays

20. How quickly did you adjust to the virtual environment experience?

☐	☐	☐	☐	☐	☐	☐
Not at all			Slowly			Less than one minute

21. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

☐	☐	☐	☐	☐	☐	☐
Not Proficient			Moderately Proficient			Very Proficient

22. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

☐	☐	☐	☐	☐	☐	☐
Not at all			Interfered Somewhat			Interfered Greatly

23. How much did the control devices interfere with the performance of assigned tasks or with other activities?

☐	☐	☐	☐	☐	☐	☐
Not at all			Interfered Somewhat			Interfered Greatly

24. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

☐	☐	☐	☐	☐	☐	☐
Not at all			Somewhat			Completely

25. How completely were your senses engaged in this experience?

☐	☐	☐	☐	☐	☐	☐
Not Engaged			Mildly Engaged			Completely Engaged

26. How easy was it to identify objects through physical interaction; like touching an object, walking over a surface, or bumping into a wall or object?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Impossible — Moderately Difficult — Very Easy

27. Were there moments during the virtual environment experience when you felt completely focussed on the task or environment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

None — Occasionally — Frequently

28. How easily did you adjust to the control devices used to interact with the virtual environment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Difficult — Moderate — Easily

29. Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not Consistent — Somewhat Consistent — Very Consistent

Appendix V – Data Collection Sheet

Subject			Date	
Unit	VERT LINAC			
Start time			End time	

Treatment Distances

F.S.D.	

Distance from inside the applicator to the skins surface. Measure in mm.

Sup-Lat	Sup-Med	Inf-Lat	Inf-Med

Fit of Setup

Was the setup correct in terms of light beam to patient marks?	YES NO

If NO what changes need to be made?

Clockwise Rotation	*Yes/No*	Anticlockwise Rotation	*Yes/No*
Couch movement Long	*Sup/Inf*	Couch movement Side	*Lat/Med*

Gantry, Bed, Collimator, Check, Pause, Double move

1	2	3	4	5	6	7	8	9	10

11	12	13	14	15	16	17	18	19	20

21	22	23	24	25	26	27	28	29	30

31	32	33	34	35	36	37	38	39	40

41	42	43	44	45	46	47	48	49	50

51	52	53	54	55	56	57	58	59	60

61	62	63	64	65	66	67	68	69	70

Upper box action, lower box time

Appendix VI – Focus Group Guide

Participant time required: 1 hour
Total focus group time: 1 hour + 10 minutes

Student perception on the use of the VERT virtual learning environment.

I. Introduction (5 min)

Welcome participants and introduce myself.

Aims of the study
This study aims to collect information about the use of VERT in radiotherapy education.
Specifically about:
- The usefulness of the system.
- Where it may be used in the syllabus.
- If it can be used as an assessment tool.

Discussion Guidelines:
I would like the discussion to be informal, so there's no need to wait for me to call on you to respond. In fact, please try and respond directly to the comments other people make. If you don't understand a question, please let us know. I am here to ask questions, listen, and make sure everyone has a chance to share.

If you aren't saying much on a topic I may interrupt you and call on you directly. If I do this, please don't feel bad about it; it's just my way of making sure I obtain everyone's perspective and opinion is included.

I do ask that we all keep each other's identities, participation and remarks private. Because of this I hope you'll feel free to speak openly and honestly.

As discussed, I will be recording the discussion and taking notes, because I want to ensure that I catch all your comments. No one outside of this room will have access to these files and they will be anonymised soon after the focus group.

Let's begin. Let's find out some more about each other by going around the room one at a time. Tell us your first name, year and main clinical site. I'll start.

Discussion (50 mins)

Hopefully you are all aware that the NHS invested money in VERT and purchased one for each HEI and for any hospital training radiotherapy students that wanted it. By now you have all used the VERT system:

Use 1	How would you describe VERT as an aid/method to learning clinical skills? Prompts: Set-up skills Equipment issues such as isocentre Patients
Use 2	What was the most important learning or insight that you gained from using VERT?
Use 3	Is there anything the staff can do to make experience better? Prompts: Group size Structure Timing
Use 4	Use 4. Is there anything that can be changed about the equipment to make the experience better?

Ass 1	How would you describe VERT as a method of assessing competence of clinical work?
Ass 2	I want to discuss both the benefits and issues you about using VERT for assessing clinical work. Firstly let's just consider what the benefits of using VERT to assess clinical work are? Prompts: Timing Standardisation X-rays/electrons Probe: How can we build on these benefits?
Ass 3	Now what are the issues of using VERT to assess clinical work? Probe: Do you think these issues could be overcome? And if so how?
Ass 4	You have used a model on a real linac and VERT to practice setups. What are the most important differences in using these two methods?

	Probe: Which method was more realistic and why?
Ass 5	What suggestions would you make to improve the way VERT could be used to assess clinical competence?

Closure (5 mins)

Though there were many different opinions about _______, it appears unanimous that _______.

Does anyone see it differently?

It seems most of you agree ______, but some think that _____.

Does anyone want to add or clarify an opinion on this?

Is there any other information regarding your experience with VERT both as a teaching tool and an assessment tool that you think would be useful for me to know?

Thank you very much for coming this afternoon/morning. Your time is very much appreciated and your comments have been very helpful.

Appendix VII – Risk Assessment

UEL Fieldwork Risk Assessment

RISK ASSESSMENT FORM FOR FIELDWORK ACTIVITY

- You are required to undertake a risk assessment prior to any fieldwork activity or research external to UEL campuses and premises

Name of Principle Investigator: David Flinton

School or Service: Cass Education

Location	**Risks identified**	**Level of risk** Low (1) Medium (2) High (3)	**Mitigating activity for each risk**	**Likely impact of mitigating activity** High (1) Medium (2) Low (3)	**Overall assessment of risk (column 3 x column 5)** Low (1) Medium (2/5) High (6/9)
VERT	Simulator sickness (Similar to motion sickness)	1	Trained observer present/ Subjects all have used the system before and are therefore aware if this affects them. Symptoms are related to time of use which is typically 1-2 hours (only 15 minutes for this study. Can be stopped by taking headset off.	1	1*
Clinical department	Crash the treatment unit into the phantom	1	Observer is a state registered radiographer who can step in to stop the movement. Students are already trained to varying degrees and have experience on the equipment	1	1

*Use of the VERT system has been ranked as low in our University risk assessment for normal use of the VERT system.

Declaration of Principal Investigator:

I confirm that I:

1) have consulted the University Risk Assessment Policy and have fully considered the risks;
2) view the relevant hazards and risks associated with the study to be at tolerable levels and within acceptable parameters;

Name of Principal Investigator: D Flinton

Signature: D Flinton

Date: 17/9/12

Appendix VIII – Participant Information Sheet & Consent Form

A Study looking at Competency based assessment using Virtual Reality.
Participant Information Sheet

I would like to invite you to take part in a research study. Before you decide whether you would like to take part it is important that you understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

What is the purpose of the study?
The study aims to look at the possibility of using the VERT system as a tool to measure a student's competence in performing electron set-ups. This research project will provide the data for my thesis that I undertaking at the Cass school of Education, University of East London (UEL)

Why have I been invited?
You have been invited to take part in this study as you are undertaking a radiography degree here at City University. All students currently on the course are being invited to take part.

Do I have to take part?
It is entirely up to you to decide whether or not to take part. If you decide not to, I will entirely respect your decision and, and you will not be penalised or disadvantaged in any way and it will not affect your grades. If you decide to take part, you are still free to withdraw from the study at any time and without giving a reason. If you do decide to take part you will be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

What do I have to do?
Firstly you will be asked to fill out a questionnaire (approx. 10 minutes) that looks at an individual's ability to experience virtual reality as real. You will then be asked to undertake two electron set-ups which will be videoed. One set-up will be done using the VERT system; the other will be performed in the clinical department on a linac. After each setup you will the complete another short questionnaire which again takes about 10 minutes each time and looks at your experience in the simulation. You will be informed which order you will be performing the two set-ups as this needs to be randomly decided. The dates and times of the two electron set-ups will then be agreed between you and the researcher. The time gap between the two set-ups will be between 1 and 3 weeks. After the set-ups a number of you will then be invited to take part in a focus group to discuss your experience and feelings about both methods which should last no longer than 1 hour.

Expenses
- You will receive travel expenses if incurred to and from the clinical site.

What do we do?
Both electron set-ups will have an observer present who will watch and video the procedure. After each set-up the observer will take measurements of the stand-off to determine the accuracy of the set-up. It is this data together with the questionnaire data that I will then analyse. The video will be analysed to determine the sequence you undertook to achieve the setup.

What are the possible disadvantages and risks of taking part?
Although I think there is almost no risk in the study, it must be remembered that the use of any machinery such as a linac carries some risk. If the observer sees any risk he will stop the set-up immediately. Also there is a small risk of side effects similar to motion sickness from the use of the virtual system, but you have all used this system before and will know if you experience these effects.

What are the possible benefits of taking part?
There will be no direct benefit to you from doing this study, other than allowing you to practice 2 electron setups in a controlled environment. It may however, improve future student assessments by allowing greater standardizing of the process, allowing it to be completed in the academic rather than clinical environment.

Will my taking part in the study be kept confidential?
All information that is collected about you during the study will be kept strictly confidential and will be stored and later destroyed in compliance with the Data Protection Act 1998. Your information will not be used or made available for any purpose other than for this research project. Transcripts of your interviews and the data from the setup and videos will be coded for anonymity and stored in a locked filing cabinet in a secure room or as an encrypted file on a password protected computer secured against unauthorised access. Anonymised data will only be made available to the supervisory team. However, the information you provide is subject to legal limitations in data confidentiality (i.e. the data may be subject to a subpoena, a freedom of information request or mandated reporting by some professions).

What will happen to results of the research study?
The study will provide the data for the research element of my Doctorate in Education. At the end of the data collection and analysis, the overall results of the research will be available on moodle. I anticipate publishing the data in scientific journals and presenting them at national and international meeting, but the identity of those who have participated will not be disclosed.

What if there is a problem?
If you would like to complain about any aspect of the study, you can contact them at researchethics@uel.ac.uk. Please include the title of the research "*Competency based assessment using VERT*." in any correspondence. UEL are the sponsors of the study. (The sponsor being the body that takes ultimate responsibility for the research and ensures that the project has adequate indemnity/insurance.)
……………………………………………………………………………………………
…
Who has reviewed the study?
This study has been formally approved by UEL's UREC (University Research Ethics Committee).

Further information and contact details
If you have any questions relating to this study please feel free to contact David Flinton.
Email: d.m.flinton@city.ac.uk
Tel: 0207 040 5688

Thank you very much for giving consideration to participating in this research project.

Please feel free to contact me with any queries.

Consent Form

Title of Project: Competency based assessment using Virtual Reality.

Name of Researcher: D Flinton.

Please initial each box.

I confirm that I have read and understand the information sheet dated 21/9/2012 (version 1.2) for the above research study and have had the opportunity to ask questions.	
I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without giving reasons and without being penalized or disadvantaged in any way.	
I understand that I will be required to complete 3 questionnaires and 2 electron setups, one on VERT and one in the department, and that I might be asked to take part in a group interview.	
I understand that any information I provide is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports, publications or presentations that will arise from the project.	
I agree to take part in the above study.	

Name of Participant ______________________ Signature ______________________ Date ______________

Name of Person taking consent ______________________ Signature ______________________ Date ______________

Appendix IX – Alderson RS-111 Phantom and approximate electron field Positions

Appendix X – Correlations & Regression Figures

Figure X.I Scatterplot of Age and Immersive tendency score (LINAC).

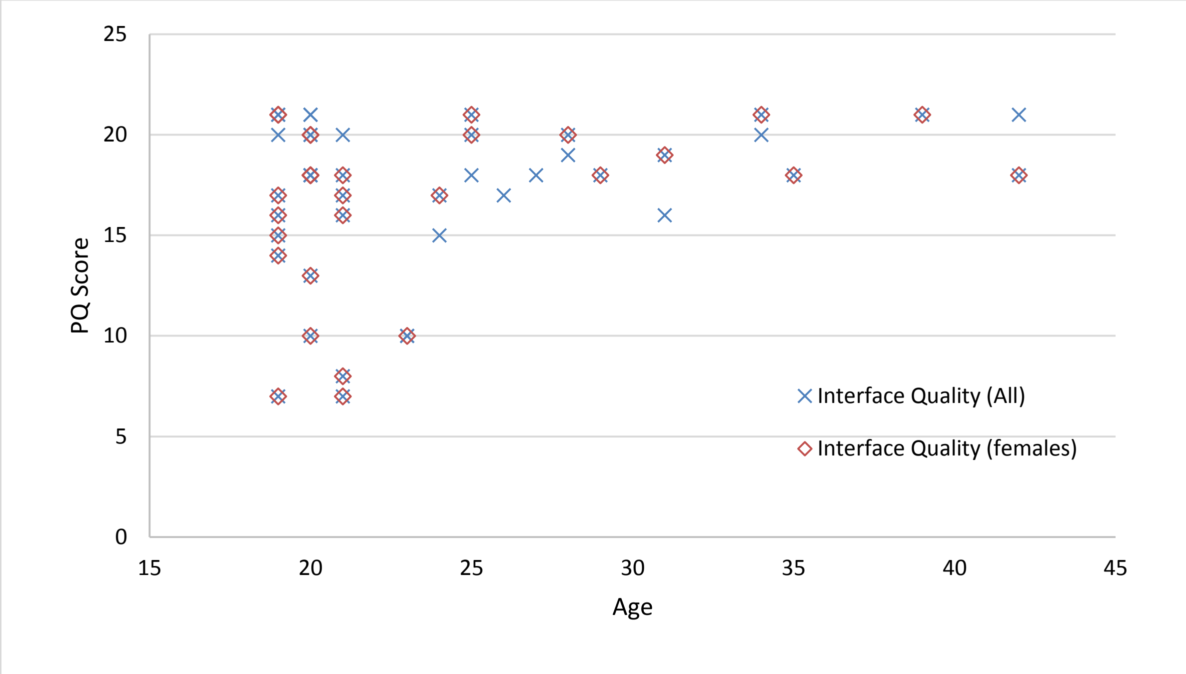


Figure X.II Scatterplot of Age and Total PQ score (males) (LINAC).

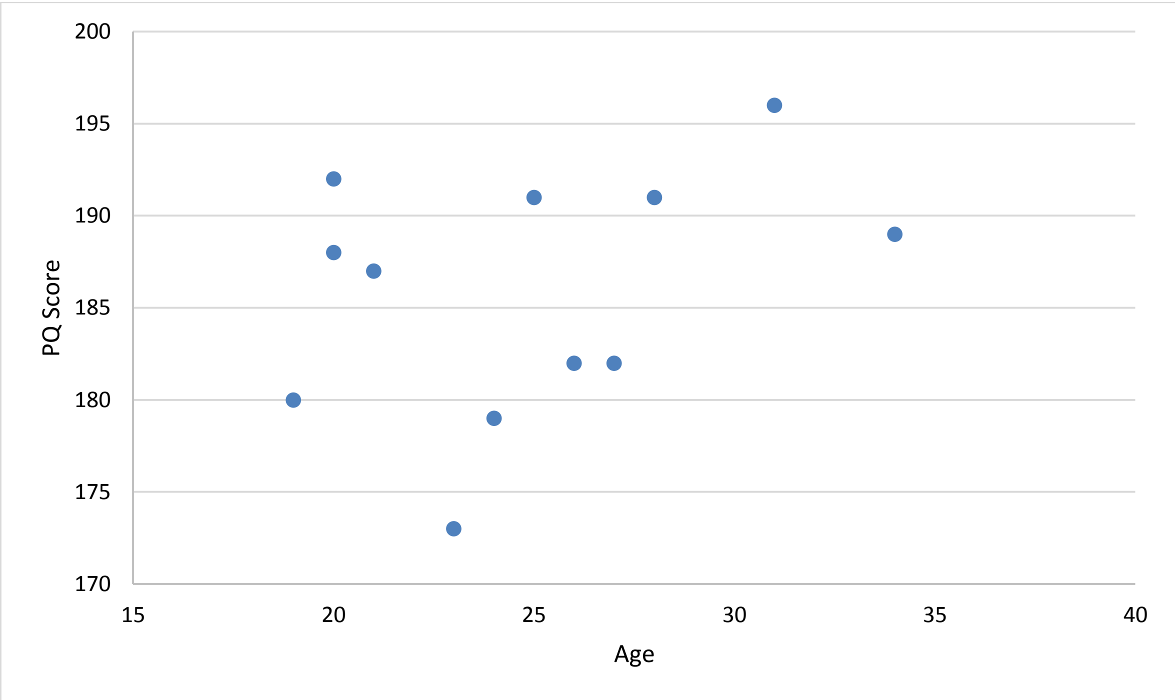
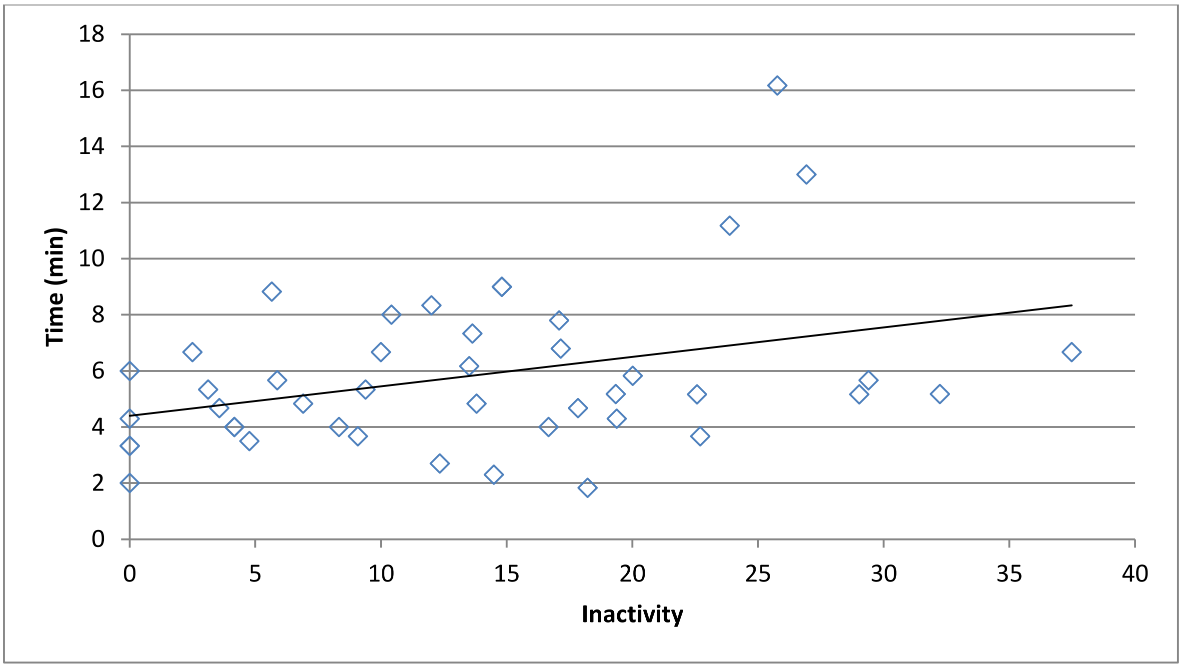
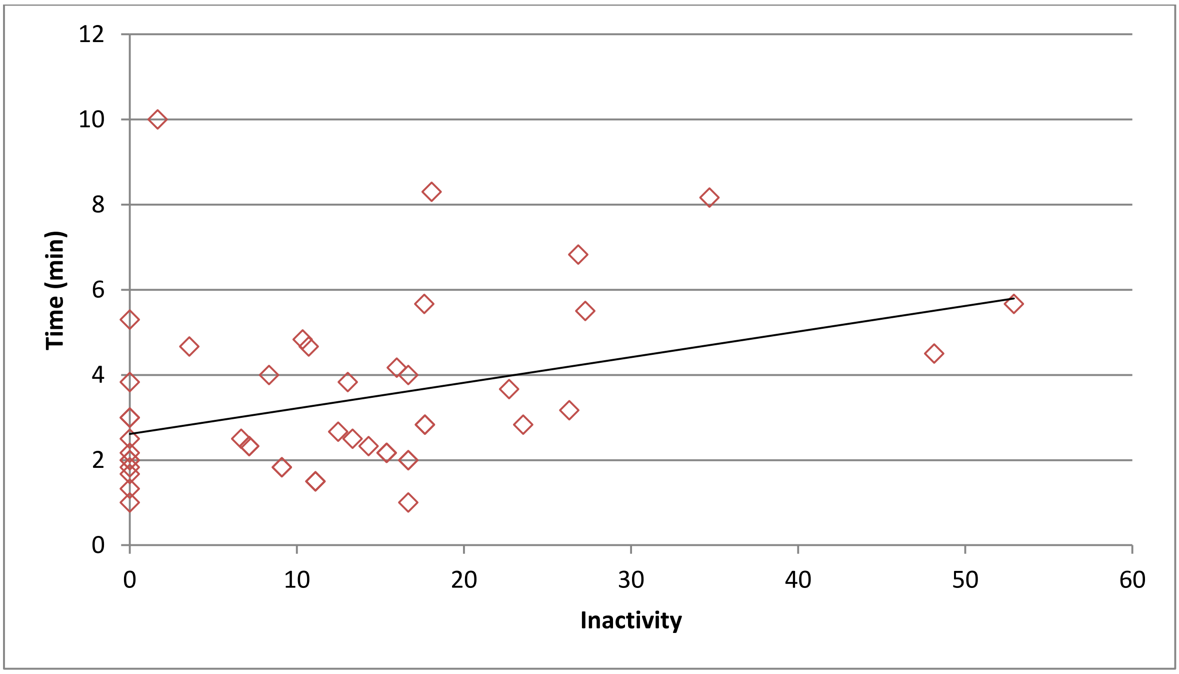


Figure X.III Relationship between Inactivity and Time. LINAC Setup 1.



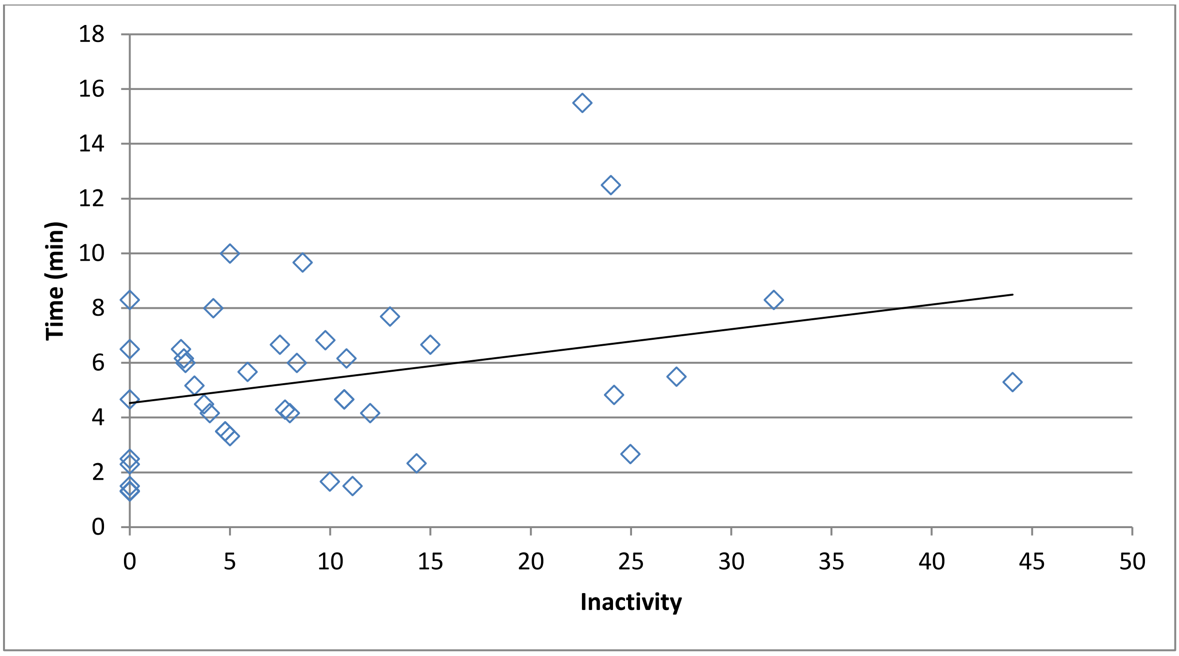
$R^2 = 0{\cdot}1383$, $p = 0{\cdot}0129$. $y = 4{\cdot}3757 + 0{\cdot}1089x$

Figure X.IV Relationship between Inactivity and Time. VERT Setup 2.



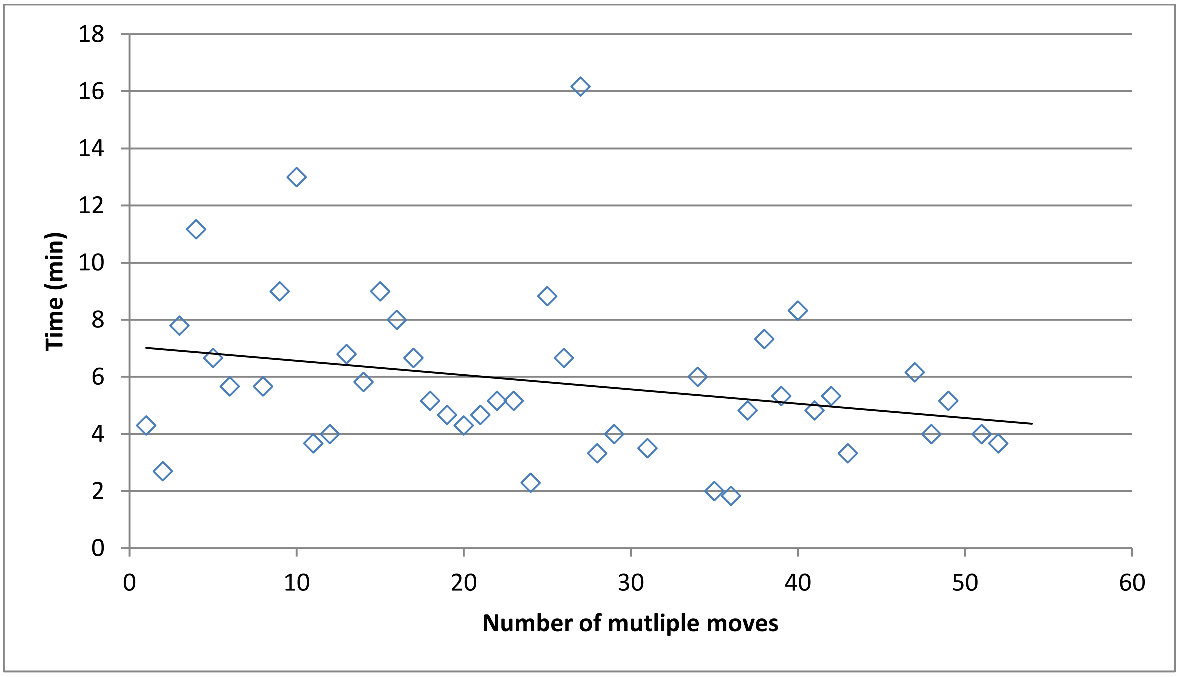
$R^2 = 0{\cdot}1367$, $p = 0{\cdot}0105$. $y = 2{\cdot}6647 + 0{\cdot}06155x$

Figure X.V Relationship between Inactivity and Time. LINAC Setup 2.



$R^2 = 0{\cdot}1062$, $p = 0{\cdot}0376$. $y = 4{\cdot}4633 + 0{\cdot}1007x$

Figure X.VI Relationship between multiple moves and Time. LINAC Setup 1.



$R^2 = 0{\cdot}1384$, $p = 0{\cdot}0129$. $y = 7{\cdot}0092 + -0{\cdot}8863x$

Appendix XI – Table of Validation Studies

Author	Simulation	Validity measured	Groups	Results
Aggarwal *et al*., (2009)	Laparoscopy, cholecystectomy 4 procedural tasks	Construct	30 inexperienced, 11 intermediate, 16 experienced	9 procedural tasks. 8 of the 9 basic skills demonstrated construct validity. Full procedure, time, $p = 0{\cdot}002$; number of movements, $p = 0{\cdot}006$; total path length, $p = 0{\cdot}033$. Procedural tasks, clipping and cutting did not differ between the 3 groups except for time taken and speed.
Bittner *et al*., (2010)	GI mentor II Endoscopy, ERCP	Face	6 novice, 6 expert	Questionnaire to all subjects. 4 questions on realism, the graphical appearance; difficulty of procedures; the overall realism of the simulation and training potential of the simulation.
		Construct		Experts outperformed novices on all metrics, time to reach papilla; Fluoroscopy time; attempts to cannulate papilla; attempts to cannulation of pancreatic duct; cannulation of common bile; contrast injection of pancreatic duct; contrast injection of common bile duct; complications. However, no metric reached significance except for the overall treatment time, $p < 0{\cdot}001$.
Bright *et al*., (2012)	TURP	Face	11 novice, 7 expert	Verbal comments to experts on completion of task about reality and appropriateness of the simulator as a learning tool.
		Construct		Time, $p < 0{\cdot}001$; amount of diathermy, $p < 0{\cdot}001$; amount of capsular resection, $p = 0{\cdot}31$.
Gavazzi *et al*., (2011)	SEP Robot, Arrow manipulation, tying a knot	Face, Content.	18 novice, 12 expert	Questionnaire to all subjects. 19 questions on usefulness of tasks, features, movement realism, depth perception and appreciation of training.
		Construct		Place arrow.

				Time, p = 0·03; dropped arrows, p = 0·005; lost arrows, p = 0·3; tool collision, NS, close entry, p = 0.02. Tying a knot. Time, p = 0·001; tightening stretch, p = 0·002; equipment dropped, p = 0·0001; winding stretch, p = 0·027; tool collision, p = 0·001.
Källström (2010)	PelvicVision simulator. Transurethral resection of the prostate	Face,	17 experts	Questionnaire asking about specific content necessary for the training of this procedure.
		Content	11 novices, 9 experts	Free text
		Construct		20 individual measures and overall result measured. Experts showed significantly better results P < 0·05 in in all parameters except for 3 (blood loss, resection volume and chip size).
Kenney *et al*., (2009)	dV-Trainer, 2 EndoWrist modules and 2 needle-driving modules.	Face, Content.	19 novices, 7 experts	Questionnaire, no details on questions, but authors commented on usefulness, incorporation into practice, acceptability and realism. Although realism and instrumentation were considered acceptable, needle driving was not.
		Construct		Time, p < 0·01; maximum force, p = 0·07; motion, p < 0·01; instrument collisions, p < 0·01; time out of view, p = 0·02; time out of centre, p < 0·01; dropped targets, p = 0·02, successful targets, p = 0·04; unattempted targets, p < 0·01, total score, p < 0·01.
Loukas *et al*., (2011)	LapVR, 3 modules Laparoscopic surgery	Construct	20 intermediate, 8 experienced	Each module was compared using 5 components, safety; technical skill; dexterity; time and velocity. The experienced group significantly (p < 0·01) outperformed the intermediate group in 14 of the 15 metrics.
McDougall *et al*., (2006)	LAPMentor Laparoscopic surgery	Face	23 novices, 24 little experience,	Questionnaire with 2 questions to novices. How useful was it to use and was it a realistic format. 94% thought it useful and 89% considered it to be realistic.

	2 modules	Content	26 some experience, 30 experts	Questionnaire with 5 questions to experts. Usefulness for training residents, 91%; better or as good as the pelvic trainer, 87%; recommendation to acquiring, 91%; appropriate testing format, 74%; used as a privileging or certifying assessment, 35%.
		Construct		9 task scores were evaluated and significant differences existed between the groups for all 9 tasks, The difference in the overall score between experts and novices was significant, p = 0·0001; as was the difference between experts and the group with little experience, p = 0·0001, but there was no significant difference between the expert and some experience group, p = 0·95)
Schreuder *et al*., (2009)	LAPSim Laparoscopic surgery	Face	15 novices, 20 intermediate, 21 experts	Questionnaire with 27 statements to all participants. 11 about the realism, 10 about the training capacities of the simulator, 5 concerning the need for training and assessment and 1 about haptic feedback. Results were presented for each group and main focus was of the publication was on the difference between the groups. Scores for the expert group were lower than the other groups. 6 of the 27 statements rated negatively and only 4 scores averaged 4 or higher on a scale of 1-5.
		Construct		20 of 82 parameters showed a significant difference between the performance of the 3 groups. Most differences were seen in basic task component where time, path length, and angular path were the most discriminative parameters. The study did not report any specific values, just that 20 parameters had a significance of 0·05 or less.
Shetty *et al*., (2012)	LAPSim Target visualisation,	Face	29 novices, 6 intermediates, 6 experts	Questionnaire to all subjects upon completion. Improved my camera handling skills. (95%); should be required for novices before assisting in the operating room

	camera navigation			(87·5%); the feedback from the program is accurate (80%); it is relevant to surgery (95%); it is a valid training tool (92·5%); it is a valid testing tool 12 (70%).
		Construct		Only looked at number of repetitions for each module and overall. Overall there was a significant difference in repetitions required to complete, $p < 0{\cdot}05$. In individual modules, coordination required more attempts for novices, $p < 0{\cdot}05$. Target visualization required more attempts for novices, $p < 0.05$. There was no statistical significance between the number of repetitions for the intermediate and the advanced groups to complete the camera navigation module.
Steinberg *et al*., (2007)	PerioSim	Content	Experts	Questionnaire looking at realism of images, realism of the mouth anatomy, usefulness of simulator instructions and use of the simulator, $p < 0{\cdot}05$.
Aggarwal *et al*., (2006)	LAPSim	Construct	7 novices, 8 intermediate, 8 experts	Statistical differences observed in total time, $p = 0{\cdot}001$; blood loss, $p = 0{\cdot}031$; and total path length $p = 0{\cdot}023$; between the three groups, experts performing better than the intermediate group which in turn performed better than the novice group.
Larsen *et al*., (2006)	LAPSim three basic skills tasks and one procedural task	Construct	10 novices, 10 intermediate, 10 experts	In the 3 basic Skills tasks, significant differences ($p < 0{\cdot}05$ were demonstrated between the expert group and the intermediate and novice group, but not between the intermediate and novice group. In the procedural task, significant differences between the experts and the less experienced groups were demonstrated in total time, blood loss and total score. The differences in the other 6 scores, instrument path length, blood pool, uterus tube cut distance, ovary diathermy damage, and unremoved dissected tissue were not significant.

Eriksen & Grantcharov (2005)	LAPSim	Construct	14 novices, 10 experts	Significant differences were demonstrated between the experienced and novice group in all seven tasks regarding time and economy of motion parameters. Significant differences existed between the groups regarding the error parameters mm tissue damage, badly placed clips, dropped clips, and blood loss. Differences in tissue damage, maximum stretch damage, failure score (rip failure, drop failure, and time-out failure,) and instrument misses did not reach statistical significance.
Woodrum et al., (2006)	LAPSim 6 tasks	Construct	9 novices, 20 intermediate, 5 experts	The 6 tasks used in this study were coordination, instrument navigation, grasping, lifting and grasping, cutting, and clip applying. Experts outperformed intermediate and novices on all tasks and metrics, $p < 0{\cdot}05$ except instrument path length and completeness of clipping where the expert group was not different to the intermediate group, but both were significantly better than the novice group.
Duffy et al., (2005)	LAPSim	Construct	10 novices, 34 intermediate, 7 experts	Less complex tasks showed discrimination, for both pass/fail and maximum level achieved, $p < 0{\cdot}03$. This task also demonstrated a difference between the times required for left instrument use, $p < 0{\cdot}05$. More complex tasks, clip applying or vessel ligation and division, were significant for pass/fail and maximum level achieved, left instrument path length, and poor clip placement, $p < 0{\cdot}02$. The complex task showed the most pronounced discrimination between groups, significant differences were noted between novice and expert groups all aspects, $p < 0{\cdot}05$). A similar significant distinction was noted when we compared novices, trainees with intermediate

Grantcharov et al., (2003)	MIST-VR 6 tasks	Construct	25 novices, 8 intermediate, 8 experts	Experienced surgeons outperformed other groups in all parameters, followed by the intermediate group and then the novices, $p < 0{\cdot}05$.
Prystowsky et al., (1999)	MIST-VR catheterisation	Content	37 novices, 14 intermediate, 9 experts	Questionnaire to novices with 6 questions. 83% of students enjoyed VR usage; 66% of students felt that VR improved their confidence; 68% indicated that they would like to use VR to learn other invasive skills. 37% felt it was harder in VR to place the IV compared to practice on fellow students.
		Construct		Comparison of pre-test and post-test success rate between groups showed no significant difference between groups in terms of performance but there was a significant difference in time.
		Concurrent		Correlation between pre-test success rate and VR success rate was low ($r = 0{\cdot}20$, NS).

Where percentages are shown this is the percentage of respondents in agreement with the statement.